RECEIVED
JAN 12 3 06 PM '00
POSTAL RATE COMMISSION
OFFICE OF THE SECRETARY

USPS–T–15

BEFORE THE
POSTAL RATE COMMISSION
WASHINGTON, D. C. 20268-0001

POSTAL RATE AND FEE CHANGES, 2000

Docket No. R2000–1

DIRECT TESTIMONY
OF
A. THOMAS BOZZO
ON BEHALF OF THE
UNITED STATES POSTAL SERVICE

Table of Contents

List of Tables

List of Figures

Autobiographical Sketch

My name is A. Thomas Bozzo. I am a Senior Economist with Christensen Associates, which is an economic research and consulting firm located in Madison, Wisconsin. My education includes a B.A. in economics and English from the University of Delaware, and a Ph.D. in economics from the University of Maryland-College Park. My major fields were econometrics and economic history, and I also took advanced coursework in industrial organization. While a graduate student, I was the teaching assistant for the graduate Econometrics II-IV classes, and taught undergraduate microeconomics and statistics. In the Fall 1995 semester, I taught monetary economics at the University of Delaware. I joined Christensen Associates as an Economist in June 1996, and was promoted to my current position in January 1997.

Much of my work at Christensen Associates has dealt with theoretical and statistical issues related to Postal Service cost methods, particularly for mail processing. During Docket No. R97–1, I worked in support of the testimonies of witnesses Degen (USPS–T–12 and USPS–RT–6) and Christensen (USPS–RT–7). Other postal projects have included econometric productivity modeling and performance measurement for Postal Service field units, estimation of standard errors of CRA inputs for the Data Quality Study, and surveys of Remote Barcode System and rural delivery volumes. I have also worked on telecommunications costing issues and on several litigation support projects. This is the first time I will have given testimony before the Postal Rate Commission.

Purpose and Scope of Testimony

My testimony is an element of the Postal Service's volume-variable cost analysis for mail processing labor. The purpose of this testimony is to present the econometric estimates of volume-variability factors used in the Postal Service's BY 1998 Cost and Revenue Analysis (CRA) for eleven "Function 1" mail processing cost pools representing operations at facilities that report data to the Management Operating Data System (MODS). I also present the economic and econometric theory underlying the Postal Service's mail processing volume-variable cost methodology. In the theoretical section, I discuss the justification for Postal Service witness Smith's application of mail processing labor volume-variability factors to non-labor mail processing costs, namely mail processing equipment costs.

Library Reference LR–I–107 contains background material for the econometric analysis reported in this testimony. It has three main parts—(1) descriptions of the computer programs used to estimate the recommended volume-variability factors, (2) descriptions of the computer programs and processing procedures used to assemble the data set used in the estimation procedures, and (3) a description of the methods used to extract MODS productivity data for use by witnesses Miller (USPS–T–24) and Yacobucci (USPS–T–25). The accompanying CD-ROM contains electronic versions of the computer programs, econometric output, and econometric input data.

I. Introduction

I.A. Overview

Clerk and mail handler costs are enormous, comprising 30 percent of the CRA total for labor costs alone, and an additional 15 percent of CRA cost in piggybacked cost components, with mail processing labor (CRA cost segment 3.1) the largest part by far. With a few relatively minor exceptions, mail processing labor costs have been assumed to be 100 percent volume-variable by the Commission. The 100 percent volume-variability assumption for mail processing, which dates back to Docket No. R71–1, has remained constant despite dramatic changes in the organization of mail processing resulting from the deployment of automation and the increasing prevalence of workshared mail. The 100 percent volume-variability assumption has been controversial, and recent rate cases have been marked by intervenor proposals to reclassify additional portions of mail processing costs as non-volume-variable.

In response to the controversies, the Postal Service produced econometric variability estimates for selected MODS and BMC cost pools (representing some 65 percent of BY96 mail processing labor costs) and revised assumptions for the remaining 35 percent in Docket No. R97–1. The Postal Service's study indicated that the degree of volume-variability varied widely among mail processing operations, and was considerably less than 100 percent overall. The the OCA, UPS, and MMA opposed the Postal Service's mail processing volume-variability study (though MMA witness Bentley did not identify any technical flaws with the study). Dow Jones and the Joint Parties sponsoring witness Higgins suppported the study.

The Commission rejected the Postal Service's Docket No. R97–1 study, finding that there was insufficient evidence to overturn the traditional 100 percent variability assumption, and citing four "disqualifying defects." However, the costing controversies that led the Postal Service to study mail processing volume-variability empirically still need to be resolved. Since Docket No. R97–1, the Data Quality Study has cast further doubt on the continued validity of the 100 percent volume-variability assumption. In this testimony, I address the defects identified by the Commission and present econometric evidence that reinforces key findings from the Postal Service's Docket No. R97–1 study. In the remaining sections of my testimony, I:

- Review the history of the analysis leading to the 100 percent assumption and the Docket No. R97–1 study (remainder of chapter I);
- Review the "disqualifying defects" of the Postal Service's R97–1 study (chapter II);
- Present a cost-theoretic framework for estimating mail processing volume-variable costs at the cost pool level (chapters III and IV);
- Present the econometric theory underlying the estimation of mail processing volume-variable costs (chapter V);
- Review the data, and data handling procedures, used for estimating mail processing volume-variable costs (chapter VI);

- Present and discuss the econometric results used in the BY98 CRA (chapter VII); and
- Discuss the status of other cost pools (chapter VIII).

I.B. Previous Postal Service research into mail processing volume-variability

This study was preceded by three major efforts to determine volume-variable costs for mail processing activities. In the late 1960s, the Post Office Department established a Cost System Task Force to develop an incremental cost analysis that was the forerunner to the present CRA. As part of its efforts to develop an incremental cost methodology, the Task Force initially attempted to estimate volume-variable costs for a variety of cost components using regression techniques. However, the Task Force determined that its statistical analysis had failed to produce a "meaningful" estimate of volume-variable costs for clerk and mail handler labor. The era of "100 percent volume-variability"[1] for the mail processing component followed, as the Task Force decided to use assumptions ("analysts' judgment"), instead of an econometric volume-variability analysis, to partition IOCS mail processing activities into 100 percent volume-variable and non-volume-variable components. The IOCS-based mail processing volume-variable cost method has survived without substantial modification since the Postal Service's inaugural rate case, Docket No. R71–1. For Docket No. R97–1,

[1] This term is a slight misnomer that is sometimes used to describe the IOCS-based mail processing volume-variability method. Several IOCS activities have always been classified non-volume-variable, but these account for a relatively small fraction of mail processing costs.

Postal Service witness Bradley presented a new set of mail processing volume-variability factors based primarily on an econometric analysis of operating data from the MODS and PIRS systems, as part of a comprehensive overhaul of the Postal Service's mail processing volume-variable cost methodology.[2] Dr. Bradley's volume-variability methods resulted in an overall volume-variable cost fraction for mail processing of 76.4 percent, versus more than 90 percent for the IOCS-based method.[3] The Commission rejected Dr. Bradley's estimated volume-variability factors in its Docket No. R97–1 Opinion and Recommended Decision. However, the Postal Service has produced FY1997 and FY1998 CRAs using its Base Year 1996 methodology from Docket No. R97–1.

I.C. Post Office Department incremental cost studies pre-R71–1

The origins of the IOCS-based mail processing volume-variable cost method, which predate the Postal Reorganization Act, had faded into obscurity as of Docket No. R97–1. The Postal Service had characterized the IOCS-based mail processing volume-variable cost method as a "convenient assumption" in its

---

[2] The other changes were to the procedures used to divide mail processing costs into cost pools for further analysis, and to the methods used to distribute volume-variable mail processing costs to the subclasses of mail. See Docket No. R97–1, USPS–T–12.

[3] See Docket No. R97–1, USPS–T–12, at 15. The volume-variable cost fraction from the IOCS-based method depends on how the clerk and mail handler costs are separated into the mail processing, window service, and administrative components. The Commission's method results in a higher volume-variable cost fraction than the Postal Service method developed in response to Order No. 1203, but the Commission's method assigns less cost to the mail processing component.

Docket No. R97–1 brief (Docket No. R97–1, Initial Brief of the United States Postal Service, at III–19), in response to which the Commission noted that the Docket No. R71–1 record contained the results of efforts to empirically estimate volume-variability factors for clerk and mail handler labor costs (PRC Op., R97–1, Vol. 1, at 68). However, it would be incorrect to say that the IOCS-based volume-variable cost method was based on the Cost System Task Force's regression analyses. Rather, the regression results to which the Commission referred convinced the Post Office Department analysts to rely on their judgment rather than statistical methods to determine clerk and mail handler volume-variable costs (see Docket No. R71–1, Chief Examiner's Initial Decision, at 20–21). In fact, the analysis that led from the regression studies to the "100 percent volume-variability" assumption covered several issues that are highly relevant to the current mail processing cost controversies. I discuss these below.

I.C.1. The analytical basis for the Postal Service's R71–1 cost methodology

The Cost System Task Force's incremental cost analysis used methods that closely resemble those underpinning the current CRA. Costs were divided into cost segments and functional components for analytical purposes, and the components were classified as volume-variable or non-volume-variable (or "fixed"). Non-volume-variable costs that could be causally traced to a class of mail or service were termed "specific-fixed"; other non-volume-variable costs were "institutional." The main difference from current CRA methods was the definition of volume-variable costs. The Task Force defined a cost component[4] as volume-variable if a percentage change in volume caused an equal percentage change in cost. In other words, for the Task Force, a volume-variable cost component was, more specifically, 100 percent volume-variable. I call this the "100 percent only" assumption.

Assuming costs must be either 100 percent volume-variable or non-volume-variable has an important function for an incremental cost analysis. It is precisely the assumption under which the incremental cost of a service is equivalent to the sum of its volume-variable cost and specific-fixed cost, a.k.a. its "attributable cost."[5] Otherwise, attributable and incremental cost differ by the "inframarginal" cost (see Docket No. R97–1, USPS–T–41, at 3–4). Under review in Docket No. R71–1, the "100 percent only" assumption was recognized as excessively restrictive by the Chief Examiner (Docket No. R71–1, Chief Examiner's Initial Decision, at 26). In no way does economic theory require volume-variability factors to be "100 percent only," because marginal cost generally varies with the level of output and may be greater than, less than, or equal to average cost. A number of cost components in the current CRA have volume-variability factors other than zero or 100 percent.

However, the Postal Service's classification of cost components as fixed or volume-variable was strongly influenced by the incremental cost study's "100

[4] I use the term "component" in the generic sense of a subdivision of CRA costs.

[5] Here, the term "attributable cost" refers to the term for the sum of volume-variable cost and "specific-fixed" cost (see USPS–LR–I–1, at H–2), as distinct from the cost concept of Section 3622(b)(3) of the Postal Reorganization Act.

percent only" assumption. Having rejected statistical analysis as a basis for the cost classification, the Task Force's experts strained to classify components as fixed or 100 percent volume-variable based on general tendencies (see Docket No. R71–1, Chief Examiner's Initial Decision, at 16–19). The Postal Service's crude division of costs was adopted because, as the Commission and the Chief Examiner agreed, no other party had presented a viable alternative (PRC Op., R71–1, at 41, 56; Docket No. R71–1, Chief Examiner's Initial Decision, at 101).

The logic of the Postal Service's cost classifications was, in many cases, extraordinarily loose. For instance, the justification of the 100 percent volume-variability assumption for the bulk of mail processing activities was that the costs "tend[ed] to be very responsive to increases in mail volume" (PRC Op., R71–1, at 4–127). There is a major lacuna between the qualitative judgment of "very responsive" and the quantification of 100 percent volume-variability. In the right context, "very responsive" could imply volume-variability factors of 60 percent or 160 percent as easily as 100 percent. Such lapses were at least equally present in the classification of costs as "fixed." The classification of window service costs as institutional, as an example, was justified by the claim that the costs "tend[ed] to vary with the growth… of population served, rather than with changes in the… volume of mail and services."[6] In retrospect, I find that these "tendencies" appear—as they do in the contemporary variability analyses for many cost

[6] Under current methodology, window service costs (Cost Segment 3.2) are nearly 50 percent volume-variable overall.

components other than 3.1—to represent cases in which the volume-variability factor is greater than zero but generally something other than 100 percent.

I.C.2. Late-1960s regression studies by the Cost System Task Force

As the Commission observed in its Docket No. R97–1 Opinion, the studies that led to the attributable cost method presented by the Postal Service in Docket No. R71–1 included efforts to use regression analysis to estimate volume-variable costs. However, these studies played no more than an illustrative role in the Postal Service's R71–1 methodology. The Cost System Task Force ultimately decided to reject its regression studies and instead use the judgment of its analysts to define volume-variable costs (PRC Op., R71–1, at 4–79 to 4–81, 4–92 to 4–102). Their decision process is relevant because the Task Force identified and discussed a number of variability measurement issues that re-emerged at the center of the Docket No. R97–1 controversies over Dr. Bradley's mail processing study. Chief among these is the need to identify and control for non-volume cost-causing factors to properly distinguish volume-variability from all other sources of cost variation. The Task Force correctly concluded that the simple regressions they ran were incapable of making these distinctions (PRC Op., R71–1, at 4–79). However, the Chief Examiner in Docket No. R71–1 concluded, and I strongly agree, that the Task Force had been too quick to dismiss the possibility of applying more sophisticated regression techniques as a remedy (Docket No. R71–1, Chief Examiner's Initial Decision, at 20–22).

The Task Force's statistical model for Cost Segment 3 was a simple regression of an index of total clerk and mail handler compensation costs on an index of mail volume, including a constant term. The Task Force estimated the regressions using annual data from FY53–FY68 (PRC Op., R71–1, at 4–107) and from FY53-FY69 (Id., at 4–125 to 4–126). In both cases, the regressions have a negative intercept and a slope slightly greater than 1. This is the entirety of the evidence that the Commission cites as an "indicat[ion] that the volume-variability of mail processing manhours was greater than 100 percent" (PRC Op., R97–1, Vol. 1, at 68). By the Commission's standards expressed in the Docket No. R97–1 Opinion, the R71–1 evidence would appear to be wholly inadequate as an empirical volume-variability study. Most of the "disqualifying" criticisms leveled by the Commission at the Docket No. R97–1 econometric models apply a fortiori to the R71–1 regressions. For instance, the authors of the R71–1 study did not attempt to collect control variables for any non-volume factors that drive cost, despite knowing that the lack of such variables likely biased their results (see below).

The Task Force's analysis did not reach the 100 percent volume-variability conclusion from the regression results. Rather, they identified a fundamental problem—apparently not accounted for in the Touche, Ross, Bailey, and Smart report mentioned by the Commission in Docket No. R97–1 (PRC Op., R97–1, Vol. 1, at 68), and also ignored in the Postal Service's R71–1 description of Cost Segment 3 —of disentangling volume from other cost-causing factors:

> <u>The underlying difficulty is that we are trying to determine the rate at which changes in volume cause costs to change, whereas changes in past costs have been due not only to volume changes, but also to changes in technology, worker efficiency, quality of service, and many other non-volume factors.</u> For example, the sharp increase in manpower costs during FY 1965 through 1968 has been attributed by the Department to not only increased volume but also, in large measure, to the adverse effect of Public Law 89-301 on productivity (PRC Op., R71–1, at 4–97, emphasis in original).

The Task Force analysis concluded that taking into account the other cost causing factors would lead to their expected cost-volume pattern, i.e., cost segments consisting of some fixed and some volume-variable cost (PRC Op., R71–1, at 4–97).

The Task Force's analysis had identified multiple regression analysis as a potential solution to the problem of disentangling the volume and non-volume drivers of clerk and mail handler cost. The Task Force legitimately cited difficulties in quantifying the non-volume explanatory factors and potential multicollinearity problems as difficulties in pursuing multiple regression as a variability measurement technique. They concluded that the basis for classifying costs as fixed or variable would "have to be analytical judgment, supported by a study of the nature of the types of work involved and whatever input and output data are available" (PRC Op., R71–1, at 4–102). Thus, the regression studies were relegated to an illustrative role in the volume-variability analysis at most. It must be recognized that many tools of econometric cost analysis that economists take for granted today, including flexible functional forms and panel data methods, were esoteric in the late-1960s and early-1970s. However, in giving up on multiple regression methods without providing so much as a correlation table showing, for instance, that problems from multicollinearity actually existed for their data, the Task Force appears to have given up too soon. Indeed, the Chief Examiner had observed that more sophisticated regression analyses had already been put to use by regulators in other industries (Docket No. R71–1, Chief Examiner's Initial Decision, at 21).

Perhaps confusing matters further, the Task Force's analysis demonstrating the inability of the simple regressions to accurately estimate volume-variability had been dropped from the cost segment descriptions presented in R71–1 (PRC Op., R71–1, at 4–125 to 4–127). Indeed, even the illustrative capability of the regressions must be judged to be extremely poor, since while the regressions might purport to demonstrate 100 percent volume-variability of total clerk and mail handler costs, the Postal Service nonetheless classified the window service and administrative components as institutional. In fact, the regressions were conducted at a level where they provide no evidence whatsoever as to the validity of the analysts' classifications of specific clerk and mail handler activities as fixed or variable. As time passed, the illustrative regressions of costs against volumes were dropped, and the descriptions of the rationale for classifying costs as fixed or variable were greatly elaborated—compare the R71–1 description (PRC Op., R71–1, at 4–127 to 4–128) with the FY96 description (Docket No. R97–1, LR-H-1, at 3–2 to 3–7)—but the quantitative evidence remained equally thin, bordering on nonexistent.

As witness Degen's testimony indicates, the traditional descriptions are especially weak on showing how costs for operation setup and material handling activities are supposed to be 100 percent volume-variable (USPS–T–16, at 5–6 et seq.). Also, the logic behind the fixed cost classifications in the traditional descriptions of mail processing was often applied inconsistently, it would seem largely due to IOCS data limitations. For instance, the FY96 description identifies some "gateway" costs (e.g., platform waiting time) as non-volume-variable but not others (e.g., portions of the collection, mail prep, and OCR operations)—a distinction that depends more on idiosyncrasies of IOCS question 18 than on operational realities (see Docket No. R97–1, LR–H–1, Section 3.1).

I.D. R97–1 Postal Service mail processing volume-variability study

I.D.1 Testimony of witness Bradley

The effort to determine the degree of volume-variability of mail processing costs returned to econometric methods with the Postal Service's study, presented in Docket No. R97–1 by Dr. Bradley. Dr. Bradley proposed new volume-variability factors for each of the cost pools defined for the Postal Service's then-new mail processing cost methodology. The volume-variability factors were derived econometrically where acceptable data were available, and based on revised volume-variability assumptions elsewhere.

Dr. Bradley used MODS data to estimate volume-variability factors for eleven Function 1 mail processing operations with piece handling data (which are updated later in this testimony), and four allied labor operations at MODS facilities. Using analogous data from the PIRS system, he estimated volume-variability factors for several BMC activities. He also estimated variabilities for remote encoding labor from remote barcode system tracking data, and for the Registry cost pool using aggregate time series data on Registered Mail volumes and costs. This portion of the analysis was a much-delayed response to a suggested refinement of the R71–1 mail processing analysis:

> Regression techniques should be applied to WLRS [Workload Reporting System, a precursor of MODS] data on manhours and piece handlings, to determine whether they will yield meaningful fixed and variable components (PRC Op., Docket No. R71–1, Vol. 4, at 4–132).

The estimated volume-variability factors were substantially lower than the IOCS-based status quo method.

The revised variability assumptions were applied to the mail processing cost pools not covered by his econometric analysis. Where possible, Dr. Bradley used econometric variabilities for similar operations as proxies. For example, a composite variability for the Function 1 Manual Letters and Manual Flats cost pools was applied to the Function 4 manual distribution (LDC 43) cost pool. For support-type activities and the cost pool defined for non-MODS mail processing, he proposed applying the system average degree of volume-variability.

I.D.2. Intervenor testimony responding to Dr. Bradley's study

Three pieces of intervenor testimony responded at length to Dr. Bradley's study. OCA witness Smith (OCA–T–600) and UPS witness Neels (UPS–T–1) opposed the adoption of the study, while Dow Jones witness Shew (DJ–T–1) favored its adoption.

Dr. Smith criticized Dr. Bradley for omitting variables, particularly wage and capital measures, that would commonly appear in economic production or cost functions. He contended that, despite statistical test results indicating the contrary, Dr. Bradley should have chosen the "pooled" regression model over the fixed-effects model in order to obtain results exhibiting the appropriate "length of run." Dr. Smith also provided a graphical analysis that purported to support the pooled regression approach. He suggested that additional analysis was required to determine the validity of Dr. Bradley's data sample selection procedures (or "scrubs") and the assumptions used to assign volume-variability factors to non-modeled cost pools. Finally, Dr. Smith claimed that Dr. Bradley's study failed to meet a set of standards for a "good" regulatory cost study.

Dr. Neels focused on Dr. Bradley's sample selection procedures, finding that Dr. Bradley had exercised his discretion to cause large reductions in sample size, with a significant effect on the regression results. Dr. Neels preferred the "between" regression model to capture the appropriate "length of run" and to mitigate potential errors-in-variables problems. However, he ultimately recommended that no econometric results should be used, claiming that the MODS workhour and piece handling data were inappropriate "proxies" for costs and volumes.

Mr. Shew approved of the Postal Service's use of extensive operational data sets from MODS and PIRS. He found Dr. Bradley's model specification to be generally adequate in its choices of output, labor input, and control variables, though he suggested that the models might be improved by incorporating data on "monetary costs" and on plant and equipment. He emphasized that Dr. Bradley's choice of the translog functional form allowed the models to exhibit a wider range of relationships between cost and outputs than simpler models, and was warranted on statistical grounds.

I.D.3. Commission analysis of the mail processing volume-variability testimony and the "disqualifying defects"

In its Opinion and Recommended Decision, the Commission commented on numerous actual or perceived flaws in Dr. Bradley's study. In rejecting Dr. Bradley's studies, the Commission highlighted four criticisms which it termed "disqualifying defects" (see PRC Op., R97–1, Vol. 1, at 65–67). To summarize:

1. The mail processing elasticities only reflect the response of costs to "volume changes that occur...within a span of only eight weeks."
2. The "scrubs" of the MODS and PIRS data are both "excessive" and "inadequate", and lead to selection biases in the elasticity estimates.
3. Some control variables assumed non-volume-variable by the Postal Service are actually volume-variable.
4. Accepting the variability estimates requires accepting a "chain of new hypotheses" regarding mail processing operations. These hypotheses include the proportionality of piece handlings and mail volumes, the non-volume-variability of Postal Service wage rates, the applicability of elasticities estimated at the sample mean to the base year and test year, and the appropriateness of pooling slope coefficients across facilities for the cost equations.

II. The Commission's "disqualifying defects" and summary of the Postal Service's response

II.A. First defect: The mail processing elasticities reflect the response of costs to "volume changes that occur... within a span of only eight weeks."

The parties in Docket No. R97–1 were nominally in agreement that the economic concept of the "long run" refers not to calendar time, but rather a hypothetical condition in which the firm is free to vary all of the factors of production. Nonetheless, the stated basis of the Commission's conclusion that Dr. Bradley's models did not reflect the appropriate "length of run" was testimony of OCA witness Smith and UPS witness Neels that focused almost exclusively on the accounting period (AP) frequency of Dr. Bradley's data and his use of a single AP lag in the models (PRC Op., R97–1, Vol. 1, at 80–81; Vol. 2, Appendix F, at 12–13).

The record in Docket No. R97–1 reflects considerable differences of opinion and some confusion over how to embody the appropriate length of run in a regression model. Much of the confusion concerned the role of Dr. Bradley's lagged Total Pieces Handled (TPH) term. Witness Neels, for example, concluded that Bradley's models were not "long run" because "they look back only a single accounting period" (Docket No. R97–1, Tr. 28/15625). On the other hand, Dr. Bradley answered several interrogatories, the apparent intent of which was to question his inclusion of even the single accounting period lag present in his regression models (Docket No. R97–1, Tr. 11/5246, 5249, 5318–23). It may be, in a sense, counterintuitive that there is any effect of lagged volume on workhours, since today's workload cannot be performed with tomorrow's labor.

In actuality, the decision to include lagged workload measures in a labor requirements model has no direct bearing on the length of run embodied in the elasticities derived from them. Rather, it is a way of incorporating the dynamics of the labor adjustment process into the model. Thus, Dr. Bradley's inclusion of a single AP lag of TPH in his model implies a labor adjustment process of approximately eight weeks. My review of witness Moden's testimony (Docket No. R97–1, USPS–T–4) and discussions with Postal Service operations experts revealed that there are two main staffing processes. One process assigns the existing complement to various operations to meet immediate processing needs, and operates on time scales on the order of hours (let alone eight weeks). However, the longer-term process of adjusting the clerk and mail handler complement operates more slowly—our operational discussions suggested up to a year. The models I present in this testimony therefore include lagged effects up to the SPLY quarter, and the volume-variability factors are calculated as the sum of the current and lagged TPH/F elasticities.

Dr. Smith's contention that the high frequency of Dr. Bradley's data, in combination with the use of the fixed-effects model, caused the Postal Service's econometric variability estimates to be "short run" was shown to be false (see Docket No. R97–1, Tr. 33/18006; USPS–T–14, at 75–77). As for the concern expressed that the horizon of the mail processing analysis reflect the "rate cycle" (see PRC Op., R97–1, Vol. 1, at 73, 79–80), real field planning processes do not take the "rate cycle" into account, so there is no operational basis for that modeling approach.

II.B. Second defect: Bradley's "scrubs" of the MODS and PIRS data are "excessive" and "ineffective" and lead to selection biases in the elasticities.

Dr. Bradley applied several sample selection criteria—or data "scrubs" as he called them—with the intent of including only the most reliable MODS data in his regressions. The Commission deemed the scrubs to be "excessive" because of the relatively large number of observations excluded as a result of applying Dr. Bradley's criteria. The Commission further concluded they were "ineffective" because the criteria cannot identify all erroneous observations in the data sets. Finally, the Commission asserted that Dr. Bradley's sample selection criteria imparted a downward bias on the elasticity estimates.

The Commission's contention in its Docket No. R97–1 Opinion that it was "evident from comparisons of estimates derived from scrubbed and unscrubbed samples that [Bradley's] scrubbing introduces a substantial selection bias that tends to depress his volume-variabilities" (PRC Op., R97–1, Vol. 1, at 84) is simply unsupported by the record in that case. Dr. Neels's own results demonstrated that there was no single direction to the changes in volume-variability factors between regressions on the full data set and Dr. Bradley's "scrubbed" data—some elasticities increased while others decreased (Docket No. R97–1, Tr. 28/15618). Joint Parties witness Higgins further showed that the effect of the "scrubs" on the estimated elasticities for the six letter and flat sorting cost pools was quite modest, and in any event trivial in comparison to the much larger omitted variables bias in the between model favored by Dr. Neels (Docket No. R97–1, Tr. 33/18018–9).

The absence of evidence that Dr. Bradley's scrubs biased his estimated elasticities was not, however, sufficient to commend their continued use in my study. I first considered whether it was necessary to employ any selection criteria beyond those absolutely required by the estimation procedures. After reviewing the relevant statistical theory, I concluded that, given the known existence of large (though sporadic) errors in the reported MODS data, employing the full "unscrubbed" data set would be inappropriate. This is because observations with extremely large errors in reported hours, Total Pieces Fed (TPF), and/or TPH can, in principle, induce large errors in the regression coefficients of any direction or magnitude. In such cases, omitting the observations, though it may appear crude, is preferable to doing nothing because it prevents biased results. Omitting the observations results only in a loss of estimation efficiency, not bias or inconsistency.

Having concluded that some selection criteria were warranted, I reviewed the details of Dr. Bradley's procedures and also considered additional procedures presented in the statistical literature. The literature considers two general classes of rules: a priori criteria, which employ independent information possessed by the researcher; and pretest criteria, in which the sample selection rules are determined by the results of a "first stage" analysis of the data. Dr. Bradley's criteria are examples of the former. The criteria are "impersonal" or "objective" (in, respectively, witness Higgins's and witness Ying's terminology; see Docket No. R97–1, Tr. 33/18014, 18149–50) in that they are applied independent of their effect on the results. An example of a pretest is an outlier detection rule that eliminates an observation from a final sample if the regression model fits the observation poorly or if the observation exerts too much "influence" on the estimates. Pretest selection procedures bring with them a significant risk of biased or inconsistent estimation (see, e.g., D. Belsley, E. Kuh, and R. Welsch *Regression Diagnostics*, John Wiley & Sons, 1980, at 15–16), which is obviously undesirable in the present context. Thus, I rejected pretest procedures as a basis for revised sample selection criteria in favor of refinements of a priori criteria similar to, but generally less restrictive than, Dr. Bradley's. I discuss these issues in detail in Section VI.D, below.

Dr. Bradley was quite candid about his belief that the large number of observations in his MODS data sets gave him latitude to impose relatively restrictive sample selection criteria. The relatively modest impact of the "scrubs" on his results (see Docket No. R97–1, Tr. 33/18019) would suggest that the restrictiveness of Dr. Bradley's sample selection criteria had little material effect on his results. Nevertheless, I determined that modifications to the procedures were warranted for two reasons. First, I have fewer observations because of the use of quarterly data over a shorter time period; second, a number of the details of Dr. Bradley's selection criteria were judgment calls that would tend to eliminate otherwise usable observations. My procedures are described in detail in Section VI.E, below. Generally, these procedures are designed to use as much of the available data as possible without admitting seriously erroneous observations. Therefore, I believe the updated sample selection criteria are not "excessive." Most of the reduction in sample size between the set of "usable" observations and my sample is required by the inclusion of additional lags of TPH/F in the models—and those observations are mostly not "discarded" per se, but rather appear as lags of included observations. I also estimated the variabilities without the sample selection procedures and found that they generally resulted in lower overall volume-variable costs for the cost pools I studied; see Appendix A.

My sample selection criteria, like Dr. Bradley's, do not and cannot identify and remove every erroneous observation. While they may appear not to address the "ineffectiveness" criticism, statistical theory indicates that the data need not be free of error for the regression results to be reliable. Rather than attempt to identify and correct for possible systematic errors—that is, errors common to all observations for a site or all sites for a given time period—in the MODS data, I control for their effects through the site-specific intercepts and flexible (quadratic) trend terms.

There is no simple method to deal with the nonsystematic, or random, error that leads to the "attenuation" phenomenon discussed by witnesses Bradley and Neels. However, theory indicates that the magnitude of the bias or inconsistency due to random measurement error increases with the measurement error variance. Or, put somewhat loosely, a process that generates relatively small (large) random errors will generate a small (large) bias. If the measurement error variance is small, the potential "harm" from the errors-in-variables problem will often be minor relative to the cost of rectifying every error. This is important because the weight conversion method for manual flats and letter operations makes it flatly impossible to rectify every error—every observation of TPH in those cost pools contains some error. TPH in the manual letter and flat distribution operations is subject to random errors in these cost pools because some of the volume of mail processed in manual operations—so-called first-handling pieces (FHP)—is measured by weighing the mail and applying an appropriate pounds-to-pieces conversion factor depending on the shape. There is always a degree of error inherent in this practice, even assuming that the conversion factors are unbiased estimates of the mean pieces per pound, due to the normal variation in the composition of mail over time and across facilities. (By contrast, piece handlings in automated and mechanized operations are generated as exact piece counts by the equipment and tend to be highly accurate.) I discuss the potential effects of measurement errors, and evidence presented in Docket No. R97–1 suggesting that any errors-in-variables effects are small, in Sections V.H and VI.E.1, below.

II.C. Third defect: Some control variables assumed non-volume-variable by the Postal Service are actually volume-variable.

In Docket No. R97–1, the Commission determined that the main control variables employed by Dr. Bradley, the "manual ratio" variable and the site-specific effects, were "likely to be volume-variable" (PRC Op., R97–1, Vol. 1, at 87; Vol. 2, Appendix F, at 39–45). In support of this conclusion, the Commission cited oral testimony of UPS witness Neels (Docket No. R97–1, Tr. 28/15795–97) in which Dr. Neels responded to questions about the potential for direct or indirect mail volume effects to explain variations in both the manual ratio and in the size of facilities. Given the circumstances, Dr. Neels's responses were unavoidably speculative, but raised legitimate issues that I investigated further.

II.C.1. Manual Ratio

Dr. Bradley interpreted the "manual ratio" variable as a parameter of the cost function that was determined largely by the mail processing technology rather than mail volumes. Clearly, technology changes can cause the manual ratio to vary without a corresponding variation in the RPW volume for any subclass—e.g., deployment of new or improved automation equipment would result in existing mail volumes being shifted from manual to automated sorting operations.

It is possible that the "manual ratio" can be affected by volume. It might be argued, for instance, that "marginal" pieces would receive relatively more manual handling than "average" pieces, and thus the manual ratio would increase, because of automation capacity constraints. However, a sustained increase in the manual ratio would be inconsistent with the Postal Service's operating plan for letter and flat sorting, which is—put briefly—to maximize the use of automated sorting operations, and, over the longer term, to deploy improved equipment that allows automated handling of increasingly large fractions of the total mail volume. Therefore, I find that to classify transient increases in the manual ratio as "volume-variable" would be to construct exactly the sort of excessively short-run volume-variability factor that the Postal Service, the Commission, the OCA, and UPS alike have claimed would be inappropriate for ratemaking purposes.

A further technical issue concerns the mathematical form of the ratio. Dr. Neels suggested that the manual ratio may be volume-variable because TPH appear in the formula. Further, the Commission showed that the derivatives of the manual ratio with respect to manual and automated piece handlings are non-zero (PRC Op., R97–1, Vol. 2, Appendix F, at 39). However, the Commission's analysis was incomplete. It can be shown that the treatment of the manual ratio does not affect the overall degree of volume-variability for the letter and flat sorting cost pools. Furthermore, if the manual ratio were to be treated as "volume-variable," its effects on the costs of individual subclasses would be small. See Appendix D for details.

II.C.2. Site-specific intercepts

The site-specific intercepts or "fixed effects," by construction, capture the effect on cost of unmeasured cost-causing factors that do not vary with volume on the margin. This is because, as the Commission correctly observed in its Docket No. R97–1 Opinion (cf. PRC Op., R97–1, Vol. 1, at 86; Vol. 2, Appendix F, at 10), the factors represented by the site-specific intercepts only capture the effect of factors that are invariant over the regression sample period. It is a logical contradiction for these factors to be both volume-variable and invariant over a sample period in which there have been significant volume changes.

Dr. Neels suspected that there could nonetheless be some "indirect" volume effect driving the persistent differences in size and other characteristics between facilities (Docket No. R97–1, Tr. 28/15796). The challenge is that the size of facilities and their mail processing operations depends not only on the volume of mail processed, but also their position in the Postal Service's network. The relevant network characteristics include both the local delivery network a facility serves and the facility's role in the processing of mail destinating elsewhere in the system. Network variables such as these are classic examples of hard-to-quantify variables that are often relatively fixed characteristics of facilities that are highly amenable to the "fixed effects" treatment. For example, a site's status as an ADC/AADC or its serving BMC are qualitative characteristics that are very unlikely to change over the near term. However, characteristics of the site's service territory are not generally fixed and can be quantified using address data for inclusion in the regression models.

My results show that the number of possible deliveries in the site's service territory is indeed an important factor in explaining persistent cost differences between sites. While possible deliveries are positively correlated with mail processing volumes—which is likely the main reason why the elasticities increase when possible deliveries are excluded from the model—they are clearly not caused by mail volumes. Rather, changes in deliveries result from general economic and demographic processes that determine household and business formation. Thus, the network effect on mail processing cost measured by the possible deliveries variable is non-volume-variable. See Sections III.B and IV.C, below, for a detailed discussion.

II.C.3. Wages

In Docket No. R97–1, the Commission counted the unknown relationship between clerk and mail handler wages and mail volumes among a series of untested assumptions underlying the Postal Service's mail processing cost methodology. In my effort to specify more standard factor demand models for the mail processing cost pools, I now include the implicit wage for the operation's Labor Distribution Code (LDC), obtained from the National Workhour Reporting System (NWRS) in the regression models. I examined the contract between the Postal Service and the American Postal Workers Union (APWU) for evidence of a direct relationship between Postal Service wage schedules and mail volumes. I found that the wage schedules in the contract depend on the employee's pay grade and length of service, but not on mail volumes.

As I discuss in Section III.C, below, it is not impossible that variations in volume could cause some variations in wages via labor mix changes, as suggested by Dr. Neels in Docket No. R97–1. The net direction of the labor mix effect of volume on wages is indeterminate. While increased overtime usage will increase implicit wages, other things equal, increased use of casual labor will decrease them. I show that per-hour compensation costs are, in fact, lower for "flexibly scheduled" labor (including overtime and casual labor) than for Straight-time hours of full-time and part-time regular clerks. However, I conclude that this type of labor mix change, and the associated decrease in wages, cannot be sustained over the rate cycle and are inappropriate to include in the volume-variability measure. The reason is that the Postal Service faces contractual restrictions that prevent it from permanently shifting its labor mix to the lowest cost labor categories, particularly casual labor. Finally, the aggregate elasticity of workhours with respect to the LDC wage is negative, as economic theory would predict. So, if the only sustainable labor mix change were one that leads to an (unobserved) increase in wages, the real labor demand would decrease in response. Based on my theoretical and empirical analysis, I determined that it is appropriate to treat the wage as effectively non-volume-variable.

II.D. Fourth defect: Accepting the variability estimates requires accepting a "chain of new hypotheses" regarding mail processing operations.

It would perhaps be more accurate to say that the MODS-based method presented by the Postal Service in Docket R97–1 shed light on assumptions that were implicit in older methods. Most of these are also untested hypotheses with respect to the Commission's method as well.

The economic assumptions underlying the MODS-based mail processing volume-variable cost methodology were subject to an extraordinary amount of scrutiny in the course of Docket No. R97–1. As a result, Postal Service witnesses Bradley, Christensen, and Degen discussed at some length assumptions of the mail processing cost methodology that previously had been implicit. Chief among these are the conditions under which the "distribution key" method for calculating (unit) volume-variable costs produces results equivalent to marginal cost, or the so-called "proportionality assumption." The proportionality assumption was, in fact, nothing new. The distribution key method was described in detail in the Summary Description of the LIOCATT-based Fiscal Year 1996 CRA, filed as LR–H–1 in Docket No. R97–1. In fact, the LIOCATT-based mail processing costs, as well as the Commission, Postal Service, and UPS methods from Docket No. R97–1, all apply IOCS-based distribution keys to MODS- and/or IOCS-based pools of volume-variable cost, and thus rely on the proportionality assumption. I argue in Section IV.E that the distribution key method is, in fact, the only feasible method to compute volume-variable costs by subclass. While the assumptions of the distribution key method are not sacrosanct, they are relatively mild and their failure would result in an approximation error, not a bias.

Dr. Bradley's mail processing elasticities (volume-variability factors) were evaluated at the sample mean values of the relevant explanatory variables, a practice used in the econometric volume-variability analyses for other cost components. The Commission expressed concern about the applicability of elasticities calculated by this method. In Section V.F, I review the relationship between the elasticity evaluation process and the goals of the costing exercise, and reconsider the "arithmetic mean" method used by Dr. Bradley along with alternative methods proposed by intervenors (for city carrier cost elasticities) in Docket No. R90–1. I conclude that the arithmetic mean method is justifiable (though cases can be made for alternative methods) and show that the results are not very sensitive to the choice of evaluation method.

It is important to note that the Commission, UPS, and LIOCATT–based mail processing volume-variable cost methods employ an additional significant untested hypothesis—the 100 percent volume-variability assumption itself. The only quantitative evidence prior to Docket No. R97–1 is more than thirty years old and was disavowed as a reliable indicator of clerk and mail handler volume-variability by its authors, as I discussed above. In Docket No. R97–1, the only statistical results that appeared to be consistent with the "100 percent variability" assumption were derived from models whose restrictions were rejected in statistical hypothesis tests. Those models were, as MPA witness Higgins put it, "'off the table' ... unworthy of consideration" (Docket No. R97–1, Tr. 33/18030). Even if the "100 percent variability" assumption were correct when originally conceived, it is, as the Data Quality Study report suggests, far from obvious that it should be equally accurate as a characterization of the volume-variability of modern mail processing operations (Data Quality Study, Summary Report, April 16, 1999, at 76).

II.E. Additional factors cited by the Commission

In reviewing the Commission's decision, I found that there were two general issues that were not explicitly stated among the "disqualifying defects," but nonetheless seemed to figure significantly in the Commission's rejection of Dr. Bradley's study. First, the Commission appeared to find the economic foundation of Dr. Bradley's regression models to be inadequate in certain respects—that his regression equations were specified ad hoc and unaccountably omitted explanatory variables that standard cost theory would consider relevant (PRC Op., R97–1, Vol. 1, at 83, 85–88). Second, the Commission seemed to find that Dr. Bradley's results defied "common sense" and that "simple, unadorned" plots of his data provided prima facie evidence in support of the "100 percent" volume-variability assumption (PRC Op., R97–1, Vol. 1, at 79; Docket No. R97–1, Tr. 28/15760).

On the first point, there is some merit to the criticisms, largely originated by Dr. Smith. I believe Dr. Bradley should have specified a more traditional labor demand function and, in particular, I find that a labor price belongs in the model. That said, though, it is not true that Bradley's models included variables that economic theory would rule out.

On the second point, the "common sense" view of mail processing needs to be re-evaluated. Mr. Degen's testimony (USPS–T–16) describes in some depth the characteristics of mail processing operations, neglected in traditional descriptions, that would be expected to lead to less-than-100 percent variabilities. As a reinforcement of the 100 percent variability assumption, "simple, unadorned" plots provide a misleading picture since they do not account for the effects of non-volume factors that may be varying along with—but are not caused by—mail volumes. Once it is agreed that a model with multiple explanatory variables is required (and this is one of the few areas of agreement among the parties from Docket No. R97–1), univariate analysis—including simple regressions and visual fitting of regression curves to scatterplots—is of no relevance.

III. Cost theory underlying mail processing volume-variable cost analysis

III.A. Cost minimization is not required to define marginal or incremental cost, but provides a useful framework for postal costing nonetheless

As part of the Postal Service's Docket No. R97–1 cost presentation, witness Panzar described the underlying cost structure in terms of an "operating plan" that need not necessarily embody the assumptions needed to define a cost function (or its "dual" production function), or to minimize costs. In no small part, Dr. Bradley gave his regression models the "cost equation" label in order to reflect the possibility, consistent with Dr. Panzar's framework, that mail processing costs are not necessarily described by a minimum cost function. Dr. Smith contended that Dr. Panzar's framework was inadequate, calling operating plans "prudent necessities of business operations" but stating that "[operating] plans and procedures do not provide the analytical form or explanatory power found in a correctly specified translog production function as defined by economists." (Docket No. R97–1, Tr. 28/15829.) Dr. Smith apparently forgot that the firm's operating plans and procedures are "real" while the economist's "production function," ubiquitous though it may be, is simply an analytical representation of those plans and procedures.[7] Whether the Postal

[7] Varian's classic textbook introduces the production function as meeting the need for "a convenient way to summarize the production possibilities of the firm" (H.R. Varian, *Microeconomic Analysis,* Second edition, W.W. Norton & Co, 1984, at 8). Chambers reminds his readers that the production function's "properties or even its existence was seriously debated" as recently as the first quarter of the twentieth century (R. Chambers, *Applied Production Analysis: A Dual Approach,* Cambridge University Press, 1988, at 6).

Service's actual plans and procedures are cost minimizing is beyond the scope of this testimony. The present analysis can be interpreted either in terms of the classical minimum cost function, or a generalized "non-minimum cost function" with a generally similar structure.[8]

The basic economic cost concepts of marginal cost and incremental cost do not depend upon cost minimization, or even upon the existence of cost or production functions, for their meaning. The Data Quality Study makes this point in a rather extreme way by arguing that economic costs are purely subjective, since as "opportunity costs" they inherently depend on the decision maker's valuation of alternative uses for resources (Data Quality Study, Technical Report #1, at 11–12). That is, the marginal or incremental (opportunity) cost is the decision maker's valuation of the resources required to produce, respectively, an additional unit or all units of a given product. I believe the important point is that any costing exercise involves a fundamentally objective exercise of measuring the "real" resource usage or demand required for some increment of a product's production, as well as a subjective exercise of valuing the resources. Observing and predicting real labor demand, which is the goal of my study, need not involve the abstract conceptual valuation problem described by the authors of the Data Quality Study. I also note that the Postal Service's costing framework wisely steers clear of the potentially extreme implications of the opportunity cost abstraction. Far from allowing "anything goes" in valuing the Postal Service's

[8] That is, the minimum cost function can be viewed as a special case of the non-minimum cost function. See Y. Toda, "Estimation of a cost function when cost is not minimized," *Review of Economics and Statistics* 58 (1976), at 259-68.

real resource usage, Dr. Panzar's framework quite reasonably values those resources at the prices the Postal Service pays for them.

III.B. "Fixed" site-specific factors, trend terms, and seasonal terms, must be held constant and are inherently non-volume-variable

In some respects, the term "volume-variability" should be self-explanatory. Cost variations not caused by volumes are not volume-variable costs. Accordingly, one of the oldest principles of postal "attributable cost" analysis is that it is necessary

> ...that certain other variables, such as productivity changes, population growth, and technological advancement, be held constant. Otherwise, it becomes exceedingly difficult to disentangle the cost-volume relationship (PRC Op., R71–1, at 48–49).

Controlling for non-volume factors, especially network effects, is central to the volume-variability analyses in other cost segments. With respect to delivery costs, statements such as, "Route time costs are essentially fixed, while access is partly variable" (R. Cohen and E. Chu, "A Measure of Scale Economies for Postal Systems", p. 5; see also LR–I–1, Sections 7.1 and 7.2) at least implicitly hold the delivery network constant. Indeed, they would likely be incorrect otherwise, since route time would not be expected to be "essentially fixed" with respect to variations in possible deliveries or other network characteristics.

The Commission's finding that the "fixed effects" are volume-variable was central among the "disqualifying defects."[9] However, the finding is based on a

---

[9] Additionally, the Commission's Opinion also appears to imply that the seasonal dummy variables and time trend, may capture volume effects. Dr. Bradley's

fundamental logical contradiction. By construction, the fixed effects capture those unobserved cost-causing factors that are constant (or fixed) over the sample period for the sites. Yet to be "volume-variable," the fixed effects would have to be responsive to changes in volume to some degree, in which case they would no longer be fixed. Additionally, the Commission also viewed the correlation between Dr. Bradley's estimated fixed effects and "volume" (specifically, the site average TPH) as causal, contending that there was no explanation other than an indirect volume effect. However, Mr. Degen describes in some detail non-volume factors that can contribute to observed high costs in high-volume operations, such as their tendency to be located at facilities in large urban areas (USPS–T–16, at 18–23).

III.C. The mail processing volume-variability analysis appropriately focuses on "real" labor demand

The mail processing variability analysis is carried out in "real" terms, that is, using workhours instead of dollar-denominated costs. The main reason for this treatment is that the rollforward model uses volume-variability factors that are free from non-volume wage effects. The rollforward process can be decomposed into the computation of "real" (constant base year dollar) costs for the test year, and adjustment of the "real" test year costs into test year dollars to account for factor price inflation. Thus, there would be some "double counting" of the inflation effect in the test year costs if the volume-variability factors used to

---

statements that those factors capture "autonomous" non-volume factors associated with the time periods are correct (PRC Op., R97-1, Vol. 1, at 88).

compute base year unit costs were to incorporate a non-volume inflation effect. Such a problem would occur if the variability analysis were carried out in "nominal" (current dollar) terms, without adequate controls for autonomous factor price inflation.

Part of the hours-versus-dollars controversy stems from the mathematical fact that variations in dollar-denominated labor "cost" can be decomposed into variations in workhours and variations in wages. However, as Dr. Bradley correctly pointed out in Docket No. R97–1, variations in wages are only of interest for the volume-variability analysis to the extent that changes in mail volumes on the margin cause them. Dr. Bradley further asserted that Postal Service wages do not respond to changes in volume, but may shift as a result of a variety of "autonomous" factors that are independent of mail volume (Docket No. R97–1, Tr. 33/17879–17889). The Commission concluded that Dr. Bradley was not necessarily wrong, but that his claims regarding the relationship between wages and volumes were unsubstantiated and required further investigation (PRC Op., Docket No. R97–1, Vol. 2, Appendix F, at 21).

I examined the wage schedules contained in the agreements between the American Postal Workers Union and the Postal Service covering the period from 1994 to the present.[10] The wage schedules do not contain any mechanism whereby volumes can directly affect wages. The agreements provide for cost-of-living and step increases in pay that depend on non-volume factors—

[10] The text of the most recent agreements has been provided as LR–I–79.

respectively, the Bureau of Labor Statistics' Consumer Price Index (specifically, CPI-W) and the length of the employee's service.

Dr. Neels also raised the possibility that volumes could affect wages indirectly by affecting the mix of workhours. However, the direction of the effect of volume on wages is ambiguous, as Dr. Neels correctly recognized:

> High-volume periods could be characterized by the more extensive use of lower-cost temporary or casual workers... It is also possible that maintenance of service standards during high-volume periods could involve greater use of overtime... pay (Docket No. R97–1, Tr. 28/15596).[11]

To examine the net effect of labor mix on wages applicable to sorting operations, I compared the average straight time wage for full-time and part-time regular clerks with the average wage for all other clerk workhours using data from the Postal Service's National Payroll Hours Summary Report (NPHSR). I use the NPHSR because it allows me to distinguish salary and benefits expenses for several clerk labor categories. Here, I separate the expenses between straight time pay of full-time and part-time regular clerks and all other clerk labor expenses. The "other" workhours category captures what can be considered "flexibly scheduled" workhours—all hours for part-time flexible, casual, and transitional clerks, plus overtime hours for regular clerks. Table 1 provides annual data for the period covered by my data set. The data clearly show that

---

[11] In addition, Dr. Neels raised the possibility that supervisory or other senior personnel might perform some mail processing work activities. This is generally beyond the scope of my testimony, since supervisor costs are part of Cost Segment 2, and my econometric analysis concerns clerk and mail handler labor costs (Cost Segment 3.1). Note, however, that the Postal Service's labor agreements generally forbid supervisors from performing "bargaining unit work" (see Article 1, Section 6 of the APWU agreement, LR–I–79).

flexibly scheduled clerk workhours are, on balance, considerably less expensive than regular clerks' straight time workhours. This phenomenon results from two main factors. First, savings in benefits costs largely offset the cost of the overtime wage premium for regular clerks. Second, salary and benefits expenses per workhour are relatively low for casual clerks, whose labor constitutes a large portion of the "flexibly scheduled" category.

Table 1. Comparison of Clerk Wages

Year	Average Straight Time Wage, Regular Clerks (salary only)	Average Straight Time Wage, Regular Clerks (salary and benefits)	Average Wage, Flexibly Scheduled Clerks and Overtime (salary only)	Average Wage, Flexibly Scheduled Clerks and Overtime (salary and benefits)
1993	$16.12	$24.79	$16.48	$19.50
1994	$16.58	$25.78	$16.09	$19.10
1995	$16.79	$26.02	$15.38	$18.70
1996	$16.94	$26.46	$15.59	$19.18
1997	$17.25	$26.91	$16.51	$20.20
1998	$17.59	$27.56	$16.78	$20.93

Source: National Payroll Hours Summary Report.

If volume peaks cause the labor mix to shift towards flexibly scheduled labor, the effect on wages would appear to be negative. Nevertheless, I do not believe that it would be appropriate to conclude that wages exhibit "negative volume-variability," or that a corresponding downward adjustment of the mail processing volume-variability factors is warranted. While Dr. Neels was correct in identifying the labor mix effects as a possible source of variation in wage rates, I believe that the labor mix effect is an excessively "short run" phenomenon. That is, while the immediate response to a change in volume may be to use flexibly scheduled labor of some kind, the Postal Service faces economic and contractual incentives to substitute towards regular workhours over the "rate cycle." In the case of regular clerks' overtime, the Postal Service is clearly capable of adjusting its complement over the course of the rate cycle, and it would be efficient to increase the complement—and straight time workhours for regular clerks—rather than systematically increase the use of overtime workhours. Witness Steele's testimony in Docket No. R97–1 shows that there are processes whereby Postal Service managers identify opportunities to employ labor in lower-cost categories (Docket No. R97–1, Tr. 33/17849–17855). However, the Postal Service's labor agreements explicitly limit its ability to sustain relatively high usage rates of labor in low-cost categories, casual labor in particular.

A central result of economics is that the real demand for a factor of production should be inverse to the factor's price. In fact, I show in Section VII.A that this result holds for mail processing labor usage—the Postal Service's staffing processes embody economic behavior in the sense that sites facing higher labor costs use less labor, other things held equal.

III.D. Relationship between volume-variability factors for labor and non-labor costs

The Postal Service's Base Year CRA applies the estimated volume-variability factors for mail processing labor cost pools to the corresponding capital cost pools (see the testimony of witness Smith, USPS-T-21, for details). While real labor input (workhours) is readily observable for use in estimating labor demand functions, capital input (as distinct from capital stocks) is not easily observable. Rather, capital input would need to be imputed from the Postal Service's fixed asset records and accounting data (which I briefly describe in Section VI.C, below). Such a process is not infeasible, but it would add an additional layer of controversy to those already present in volume-variability estimation for labor costs. Deploying reasonable assumptions to link labor and capital variabilities is a simple, feasible alternative.

In fact, the capital and labor variabilities will be identical, in equilibrium, under the assumption that the cost pool-level production (or cost) functions are *homothetic.* Homotheticity implies that changing the level of output of the operation will not alter relative factor demands such as the capital/labor ratio, in equilibrium (and other things equal). In the empirical factor demand studies, this assumption has been used to allow the constant returns to scale assumption to be tested in a dynamic system.[12] Intuitively, in an automated sorting operation, the possibilities for increasing output by adding labor without increasing capital input via increased machine utilization are limited; adding machines without labor to run them would be similarly futile. Thus, if a one percent increase in output

---

[12] See, e.g., M. Nadiri and I. Prucha, "Dynamic Factor Demand Models, Productivity Measurement, and Rates of Return: Theory and an Empirical Application to the US Bell System," *Structural Change and Economic Dynamics*, Vol. 1, No. 2 (1990) at 263–289. Interestingly, Nadiri and Prucha found equilibrium output elasticities of capital and labor for the Bell System of approximately 0.64, with labor input tending to be more elastic and capital input less elastic in the short run.

(piece handlings) in an operation led to an *X* percent increase in real labor input, where *X* is the degree of volume-variability for labor input, it would also lead to an *X* percent increase in real capital input. This implies that the equilibrium labor and capital variabilities are identical.

IV. Economic modeling of mail processing labor cost

IV.A. Volume-variability factors can be obtained from labor demand functions defined at the mail processing operation (cost pool) level

The Commission noted in its Docket No. R97–1 Opinion that Dr. Bradley's characterization of his mail processing models as "cost equations" having an undefined relationship to standard economic cost theory had caused confusion among the parties as to the meaning of his results (see PRC Op., Docket No. R97–1, Appendix F, at 7–8). In my opinion, much of the confusion should have been resolved once the "cost equations" were interpreted more conventionally as labor demand functions (see PRC Op., Docket No. R97–1, Vol. 1, at 83). Economic cost theory provides the powerful result that cost, production, and factor supply or demand functions all embody the same information about the underlying production process. Therefore, estimating labor demand functions, rather than cost or production functions, to obtain the volume-variability factors is a theoretically valid modeling approach.[13]

I agree with the Commission's conclusion in Docket No. R97–1 that organizing mail processing costs by operational cost pools "clarifies subclass cost responsibility" (PRC Op., Docket No. R97–1, Vol. 1, at 134). In my opinion, defining the mail processing production processes at the operation (cost pool)

[13] For a comprehensive treatment of the relevant theory, see R. Chambers, *Applied Production Analysis: A Dual Approach*, Cambridge University Press, New York, 1988.

level, rather than at the facility level, greatly facilitates the economic analysis of sorting operations' costs for both the volume-variability and distribution steps.

For the sorting operations, the main advantage of using cost pools as the unit of analysis is that the cost pools can be defined such that they represent distinct (intermediate) production processes with separate, identifiable, and relatively homogeneous, inputs (e.g., labor services) and outputs (processed pieces, or TPF).[14] That is, an individual clerk cannot simultaneously sort mail at a manual case and load or sweep a piece of automation equipment, nor is the ability to process mail in one operation contingent on another operation being staffed.

Certain other mail processing operations, particularly mail processing support and allied labor operations, would be expected to exhibit some form of joint production, as Dr. Smith indicated (Docket No. R97–1, Tr. 28/15830–15831). Both mail processing support and allied operations can be characterized as having multiple outputs—for example, in the form different types of "item" and container handlings, or support of several "direct" mail processing operations—that are produced using a common pool of labor resources. Thus, the economic models underlying the analysis of labor costs in allied and mail processing support operations should be distinct from those applicable to distribution operations—just as Dr. Bradley's allied labor models were distinct from his models of distribution operations. Such economic distinctions are most easily

[14] Technically, the cost pools are "nonjoint in inputs," which allows the multioutput technology to be represented as a set of conventional production functions. See Chambers, 1988, p. 287.

made when operational knowledge of the cost pools is combined with economic theory.

IV.B. Cost theory and selection of variables

In Docket No. R97–1, the Postal Service and other parties agreed on the general point that there are many explanatory factors that must be taken into account to accurately estimate volume-variability factors for mail processing operations. Indeed, Dr. Bradley was criticized for not including a sufficiently broad set of control variables in his regression models (see, e.g., PRC Op., R97–1, Vol. 1, at 85). OCA witness Smith specifically claimed that Dr. Bradley should have included measures of wages and capital in his regression equations, citing the textbook formulation of the cost function. Dr. Smith also made the contradictory arguments that, despite the need to control for a number of potential cost-causing factors, it was theoretically inappropriate for Dr. Bradley to include any explanatory variables in his models other than output (i.e., TPH), wages, capital, and a time trend (Docket No. R97–1, Tr. 33/18078–9). The Commission largely concurred with Dr. Smith's criticisms (PRC Op., R97–1, Vol. 1, at 80–83, 85–88; Vol. 2, at 1–2, 8, 12–13).

As a general matter, I find that Dr. Bradley's lack of stated cost theoretic underpinnings for his mail processing study added unnecessary confusion to the Docket No. R97–1 proceedings. However, the effects of the confusion are largely cosmetic. For example, once it becomes clear that Dr. Bradley's "cost equations" are more properly interpreted as labor demand functions (PRC Op., R97–1, Vol. 1, at 83), it should be equally clear that the elasticity of labor demand with respect to output is the appropriate economic quantity corresponding to the ratemaking concept of the "volume-variability factor." At the same time, the labor demand function interpretation of the models points out some potentially substantive ways in which Dr. Bradley sidestepped orthodox economic cost theory in his mail processing analysis. Dr. Smith is correct that certain economic variables, such as the wage rate, would normally be included in either a labor cost or labor demand function. Indeed, a Postal Service interrogatory to Dr. Smith seemed to be intended to point him in this direction (see Docket No. R97–1, Tr. 28/15909). As the Commission observed, even the "operating plan" framework described by witness Panzar assumed that the Postal Service's behavior would be "economic" in the sense that the "plan" would generally depend on factor prices (see Docket No. R97–1, USPS–T–11, at 14–15). I am in full agreement with Dr. Smith and the Commission that, to the extent data are available, additional variables indicated by economic theory should be constructed and included in the regression models.

However, textbook economic theory cannot specify the *full* set of relevant cost causing factors for any particular applied study. To create an adequate econometric model, it is necessary to identify the factors that sufficiently bridge the gap between generic theory and operational reality. This requires expert knowledge specific to the system under study. Therefore, I also agree with Postal Service witness Ying that Dr. Smith was in error to suggest that generic cost theory can be used to exclude factors that actually affect costs from the regression models (Docket No. R97–1, Tr. 33/18144). From a theoretical perspective, any factors that affect the amount of inputs needed to produce a given output will appear in the production function—and thus the derived cost and/or labor demand functions. In fact, as Dr. Ying indicated, most of the recent literature on applied cost modeling uses cost functions augmented with variables to reflect technological conditions, which are more general than the generic textbook specification referenced by Dr. Smith (Docket No. R97–1, Tr. 33/18144). The general cost functions used in the applied econometrics literature allow for network variables, other control factors, and time- and firm-specific shifts in the cost structure.[15]

From a statistical standpoint, it is well known that omitting relevant explanatory variables from a regression model generally leads to bias. In the cost estimation literature, the result that estimates of cost and/or factor demand function parameters will be biased unless all relevant "technological factors" are taken into account dates back at least to a 1978 paper by McFadden.[16] Specifically, there is no theoretical or statistical justification for excluding the "manual ratio" variable, which, as a measure of the degree of automation, is clearly an indicator of the sites' organization of mailflows in letter and flat sorting

---

[15] See, e.g., L. Christensen, D. Caves, and M. Tretheway, "Economies of Density Versus Economies of Scale: Why Trunk and Local Service Airlines Differ," *Rand Journal of Economics*, Winter 1984, at 471.

[16] D. McFadden, "Cost, Revenue, and Profit Functions," in M. Fuss and D. McFadden, eds., *Production Economics*, Amsterdam: North-Holland Press (1978). The underlying statistical theory of omitted variables bias is considerably older.

operations. To exclude the "manual ratio," or, indeed, any other variable that actually explains costs, is to introduce the potential for omitted variables bias to the results.

IV.C. Two principal "cost drivers:" mail processing volumes and network characteristics

My discussions with Postal Service operations experts indicated that both volumes and network characteristics are important factors that drive the costs in the sorting operations. The relevant network characteristics potentially include an operation's position in the overall mail processing network, mail flows within a site, and characteristics of the site's serving territory. These factors, often in conjunction with volumes, determine the length of processing windows, the complexity of mail processing schemes, the relative amount of labor required for setup and take-down activities, the operation's role as a "gateway" or "backstop", and other indicators of the level of costs and the degree of volume-variability. Earlier Postal Service studies have also identified the combination of volume and network characteristics—particularly, characteristics of the local delivery network—as drivers of mail processing space and equipment needs (see "Does Automation Drive Space Needs?" Docket No. R90–1, LR–F–333).

Volume and network characteristics interact in complicated ways, but volume does not cause network characteristics. Recipients (addresses) must exist before there is any need to generate a mail piece.[17] Witnesses Degen and Kingsley discuss the operational details more fully, but I feel it is worth highlighting a few examples here. Relatively short processing windows would tend to require schemes be run on more machines concurrently, and hence require more setup and takedown time, to work a given volume of mail. The number of separations can influence the number of batches by requiring that certain schemes be run on multiple pieces of equipment. Volume increases may require additional handling of trays, pallets, and rolling containers, but to some extent they will simply lead to more mail in "existing" containers—this is an important way in which economies of density arise in mail processing operations. The impact of these non-volume factors is not limited to automated sorting operations. The average productivity of manual sorting operations would be expected to be lower, the more complicated the sort schemes. It would not be unusual at all for clerks to be able to sustain much higher productivity levels sorting to relatively small numbers of separations than would be attainable by their colleagues working more complex schemes. Such systematic productivity differences are clearly not driven by volume, but rather by non-volume network characteristics.

Modeling network characteristics is inherently challenging. There is no particular difficulty in counting network nodes or other physical characteristics. However, the details of the network's interconnections tend to be difficult if not

[17] The substantial variation in handlings (TPH) per delivery point illustrates the lack of causality. For Q4 FY 1998 TPH per delivery point averaged 888 with an interquartile range of 412.

impossible to quantify. I expect that there will be considerable variation in these hard-to-quantify characteristics between sites, but—after accounting for the quantifiable characteristics—generally little variation over time for any specific site. For example, the (easy to quantify) number of possible deliveries in a site's service territory will tend to vary more-or-less continuously, but the geographical dispersal of its stations and branches will not. I used essentially the method of Caves, Christensen, and Tretheway (1984)[18] of including in the regression models available quantitative variables pertaining to network characteristics in a flexible functional form in conjunction with site-specific qualitative (dummy) variables or "fixed effects" to capture non-quantified network characteristics.[19]

I initially considered three quantitative variables related to the site's serving territory: the number of possible deliveries (served by the REGPO), the number of 5-digit ZIP Codes, and the number of post offices, stations and branches. I found that the hypothesis that the coefficients on these variables were jointly zero could be rejected for all operations. However, I found evidence that the ZIP Code and office variables were poorly conditioned because of high correlation with possible deliveries and little variation within sites. Thus, my preferred specification employs only possible deliveries.

---

[18] D. Caves, L. Christensen, M. Tretheway, "Economies of Density Versus Economies of Scale: Why Trunk and Local Service Airlines Differ," *Rand Journal of Economics*, Winter 1984.

[19] As I discuss in Sections V.D and V.H, capturing the effect of unobserved network characteristics, while important, is not the only reason to allow for site-specific shifts in the labor demand functions.

Like Dr. Bradley, I include the "manual ratio"—the fraction of all piece handlings for the shape of mail processed manually—in the labor demand function. I discuss the Commission's conclusion that the manual ratio is "volume-variable" in Section IV.F, below. I note that the manual ratio can be viewed as a control variable capturing the organizaiton of local mail flows, as well as an indicator of the "hygiene" of an operation's mail (as in Dr. Bradley's interpretation).

IV.D. In MODS sorting operations Total Pieces Fed (TPF) is the appropriate measure of mail processing volumes

An economic analysis of a production process requires that an output (or outputs) of the process be identified. For mail sorting operations, the "outputs" are the sorted pieces handled therein ("piece handlings" for short). That piece handlings constitute the output of sorting operations was, in fact, a relatively rare point of agreement between Dr. Bradley (see Docket No. R97–1, USPS–T–14, at 6) and Dr. Smith (who was critical of the lack of an explicit theoretical framework and believed additional variables were needed; see Docket No. R97–1, Tr. 28/15825–31).

Most of the workload measurement effort in MODS is, in fact, geared to measuring volumes of mail handled in sorting operations. The system offers three candidate volume measures, First Handling Pieces (FHP), Total Pieces Handled (TPH), and Total Pieces Fed (TPF). The FHP measure has two conceptual deficiencies. First, as its name suggests, an FHP count is only recorded in the operation where a piece receives its first distribution handling within a plant. A piece that is sorted in both an OCR and a BCS operation would be part of the output of both operations, but no FHP would be recorded in the downstream operations. Second, the work content per FHP may vary widely from piece to piece even within an operation because some mailpieces—e.g., nonpresorted pieces, and pieces addressed to residences (as opposed to post office boxes)—require more sorting than others.

The TPH measure is conceptually superior to FHP as an output measure for sorting operations because a TPH is recorded in every operation where a piece is successfully sorted, and a piece that requires multiple sorts in an operation generates multiple TPH. A further advantage of TPH (and TPF) is that it is based on actual machine counts, rather than weight conversions, for automated and mechanized sorting operations. Therefore, TPH and TPF data for automated and mechanized sorting operations are not subject to error from FHP weight conversions.[20] However, for automated and mechanized operations, TPH excludes handlings of pieces not successfully sorted ("rejects"), so it does not quite capture these operations' entire output. Therefore, I use TPF, which includes rejects as well as successfully sorted pieces, as the output measure for

---

[20] This important point caused considerable, if needless, confusion in Docket No. R97–1. Dr. Neels erroneously claimed that TPH in automated and mechanized operations were subject to FHP measurement error based on the mistaken belief that those TPH are computed as the sum of FHP derived from weight conversions and subsequent handling pieces (SHP). Dr. Neels cited the definition in Section 212.211 of the M–32 MODS handbook in support of his claim (see Docket No. R97–1, Tr. 28/15602; USPS–LR–H–147). However, sections 212.222 and 212.223 (on the same page of the manual) clearly state that the TPH and TPF data for automated and mechanized operation are obtained from machine meter readings or end-of-run reports. For these operations, TPH and FHP measurement are independent.

automated and mechanized sorting operations (BCS, OCR, FSM, LSM, and SPBS). Separate TPH and TPF are not recorded for manual operations since those operations do not generate rejects, and I therefore use TPH as the output measure for the remaining operations.

In Docket No. R97–1, UPS witness Neels also claimed that the TPH data used by Dr. Bradley were an inadequate "proxy" for "volumes" in the mail processing model (Docket No. R97–1, Tr. 28/15999–16000). It must be noted that the validity of Dr. Neels concerns have absolutely no bearing on the need to estimate elasticities with respect to piece handlings. Strictly speaking, Dr. Neels's "proxy" criticism describes certain ways in which the assumptions of the distribution of volume-variable costs to subclasses could potentially fail to hold. Recall that Postal Service witness Christensen pointed out in Docket No. R97–1 that the "distribution key" method used by the Postal Service breaks down the connection between cost and volume into a two-step procedure. The first ("attribution") step requires measurement of the elasticity of an operation's costs with respect to its outputs (or "cost drivers"); the second ("distribution") step requires estimates of the elasticities of the cost drivers with respect to subclass (RPW) volumes. However, since it is impossible to estimate the latter elasticities, given the large number of subclasses for which volume-variable costs are computed and the low frequency of RPW time series data, the distribution step proceeds under simplifying assumptions in the "distribution key" method. (I discuss the implications of the distribution key method further in the next section). In Docket No. R97–1, Dr. Bradley carried out the first step, whereas Mr. Degen handled the second step (Docket No. R97–1, Tr. 34/18222–3). Dr. Neels's criticism was actually misdirected—it should have been directed at the mail processing cost distribution study rather than to the volume-variability study.

IV.E. The "distribution key" method is the only feasible way to compute mail processing volume-variable costs by subclass; its underlying assumptions are minimally restrictive as applied by the Postal Service

Directly estimating the elasticities of cost drivers with respect to RPW volumes is infeasible, so the CRA extensively uses the "distribution key" method to compute volume-variable costs by subclass. The "distribution key" method uses shares of the cost driver by subclass to distribute the pool of volume-variable costs from the "attribution step." The cost driver shares for a mail sorting cost pool can be estimated by sampling the pieces handled in the operation. In the case of mail processing operations, the sample is the set of IOCS "handling mail" tallies. The computational advantage of the distribution key method is that it dispenses with the marginal analysis of the relationship between volumes and the driver. The price of simplicity is what has been termed the "proportionality assumption." Formally, the distribution key method and the constructed marginal cost method are equivalent when the cost driver is a linear function of the mail volumes or, equivalently, the number of handlings of a representative piece of a given subclass is "constant."

There is no inherent bias in the proportionality assumption. To the extent the assumption does not hold, all that arises is an approximation error from using a linear function relating volumes and cost drivers to stand in for the true nonlinear relationship. It is also important not to read too much into the assumption that the proportions are constant. In this context, the "constancy" of handlings per piece does not mean that every piece of a subclass has the same work content. Indeed, all subclasses involve some averaging of work content over origin/destination pairs and other characteristics of individual pieces. Rather, it amounts to a limited assumption of reproducibility—holding other things equal, two otherwise identical pieces will follow a materially identical processing path. For example, I expect that my remittance to a non-local credit card issuer (sent via First-Class Mail) will require more BCS sorts to reach its destination than my payment to the local electric utility. But, other things equal, I expect next month's credit card payment to require the same number of sorts as this month's. If there happened to be a change in the processing pattern, it would likely be due to some factor other than sending in the additional piece for the next month's payment.

The Postal Service's methods recognize that the absolute and relative amount of handlings per piece may vary over time, due to changes in Postal Service operations, mailer behavior, or other factors. The annual updates of the cost pool totals and distribution key shares permit the assumed handling levels and proportions to vary over time. Indeed, if it could be assumed that processing patterns and subclass characteristics were stable over a multi-year period of time, it would be possible to pool multiple years' IOCS data to improve the statistical efficiency of the distribution keys. The assumption implicit in the Postal Service's method that major changes in operations will not take the form of drastic intra-year changes is not very restrictive, given that most national deployments of new equipment and substantial changes to operations require years to complete. Likewise, it is hard to envision rapid and drastic changes in the average work content of the mail subclasses in the absence of correspondingly drastic changes to worksharing discounts and other economic incentives facing mailers. Of course, to the extent such changes were anticipated between the base year and test year, it would be appropriate to include a corresponding cost adjustment in the rollforward model.

Dr. Neels correctly observed that failure of the proportionality assumption does not impart a bias in any obvious direction (Docket No. R97–1, Tr. 28/15599, at 2–6). As a result, Dr. Neels's suggestion that "[c]hanges in the relationship between piece handlings and volume could mask significant diseconomies of scale" (Id., at 12–13) relies on flawed logic. To illustrate the point, he suggests that an increase in volume could lead to "increases in error sorting rates [sic]" (Id., at 15). The logical error is that Dr. Neels's illustration transparently violates the ceteris paribus principle since it presupposes a change in mailpiece characteristics such that the marginal piece (of some subclass) would be less automation-compatible than the average piece, in addition to the change in volume. As a practical matter, the example seems all the more off the mark given the Postal Service's ongoing efforts to improve the functionality of its automation equipment and to ensure the automation-compatibility of automation-rate mail.

Finally, Dr. Neels's criticism applies equally to all of the mail processing volume-variable cost distribution methods, including (but not limited to) the UPS method proposed in Docket No. R97–1, the PRC method adopted in Docket No. R97–1, and the LIOCATT-based method in place prior to Docket No. R97–1. Insofar as the distribution key method is universally used, has no feasible alternative, and imparts no obvious bias on the measured volume-variable costs, I find that Dr. Neels raised some potentially interesting issues but did not provide a constructive criticism of the available costing methods.

IV.F. The manual ratio should be treated as non-volume-variable

Dr. Bradley interpreted the "manual ratio" variable as a parameter of the cost function that was determined largely by the mail processing technology rather than mail volumes. Clearly, technology changes can cause the manual ratio to vary without a corresponding variation in the RPW volume of any subclass. For instance, deployment of the Remote Barcode System has allowed the Postal Service to shift mail volumes that formerly required manual processing because of lack of a mailer-applied barcode or OCR-readable address to automated sorting operations.

In some circumstances, the "manual ratio" might be affected by volume. It could be argued that "marginal" pieces of mail would receive relatively more manual handling than "average" pieces because of automation capacity constraints, so a volume increase would tend to increase the manual ratio. However, a volume effect on the manual ratio that is contingent on automation capacity limitations is short-run by definition. To the extent that the Postal Service can potentially adjust its automation capacity over the course of the "rate cycle" to allow marginal volumes to be processed on automation (consistent with its operating plan) the volume effect on the manual ratio would be "excessively short-run." Thus, to classify the manual ratio as "volume-variable" for that reason would be to construct the sort of overly short-run volume-variability analysis that the Postal Service, the Commission, the OCA, and UPS alike have claimed would be inappropriate for ratemaking purposes.

A technical issue for the treatment of the manual ratio variables concerns the mathematical form of the ratio. Dr. Neels suggested that the manual ratio may be volume-variable because TPH appear in the formula. While it is true that the manual ratio depends on TPH (both automated and manual), that does not establish the "degree of volume-variability" for the manual ratio. The Commission showed that the derivatives of the manual ratio with respect to manual and automated piece handlings are nonzero but also that they have opposite signs (PRC Op., R97–1, Vol. 2, Appendix F, at 39). Since a volume change will normally cause changes in both manual and automated handlings, the manual ratio effects at the subclass level will partly cancel out. The canceling effect will be greater to the extent a subclass is responsible for a similar share of handlings in both the manual and automated operations for a given shape. Furthermore, when summed over all subclasses (by cost pool), the manual ratio effects cancel out. Thus, the overall degree of volume-variability for the letter and flat sorting cost pools does not depend on whether or not the manual ratio is treated as “volume-variable.” Details of the supporting calculations are provided in Appendix D to this testimony.

V. Econometric modeling of mail processing labor cost

V.A. Volume-variability factors cannot be intuited from simple plots of the data

During the hearings on the Postal Service's direct case in Docket No. R97–1, Chairman Gleiman asked Dr. Bradley to confirm the intuition

> ...that if costs vary 100 percent with volume, the graph of those costs and the volume data points should resemble a straight line with a 1-to-1 slope (Docket No. R97–1, Tr. 11/5578, at 4–6).

Dr. Bradley agreed, and even added that the line should go through the origin (Id., at 8–9; 11).[21] In my opinion, Dr. Bradley should not have confirmed Chairman Gleiman's intuition. It has been understood since Docket No. R71–1 that to measure "volume-variability," it is necessary to hold constant the non-volume factors that affect costs. By virtue of its lack of additional "control" variables, a simple regression (or plot) of cost on volume cannot do so—it is subject to omitted variables bias. Dr. Bradley indicated as much in his response to Chairman Gleiman's subsequent question asking whether Dr. Bradley had plotted the "cost-volume relationship" for the modeled operations. Explaining why he had not plotted the relationship, Dr. Bradley stated:

> The cost-volume relationship you talk about is what's known as a bivariant [sic] analysis, and it doesn't account for the variety of other factors which are changing as those two things change (Id., at 12–18).

[21] Dr. Bradley's statement that the line should additionally pass through the origin was in error. As a general matter, the cost surface passing through the origin is neither necessary nor sufficient for the 100 percent volume-variability result.

In effect, Dr. Bradley did not produce the data plots because they were irrelevant and misleading with respect to the goal of obtaining unbiased (or consistent) estimates of the elasticities.

Despite the fundamental inadequacy of simple cost-volume plots as a statistical tool, both Dr. Neels and Dr. Smith offered interpretations of cost-volume plots in support of the 100 percent volume-variability assumption. Dr. Neels found the plots to be "visually compelling" evidence of 100 percent volume-variability (Docket No. R97–1, Tr. 28/15847).[22] Visual inspection of plots of hours against TPH constituted the entirety of Dr. Smith's quantitative analysis (Docket No. R97–1, Tr. 28/15826–15854), despite his claim that Dr. Bradley's models were underspecified (Docket No. R97–1, Tr. 28/115826–15831). Indeed, Dr. Smith's quantitative and qualitative analyses were seriously at odds, since the former was subject to the criticisms in the latter. Mr. Higgins observed that Dr. Smith's visual analysis did not (indeed, could not) take into account the additional variables one might expect to find in a cost or factor demand function, and was therefore subject to omitted variables bias (Docket No. R97–1, Tr. 33/17993–4).

In addition to conceptual shortcomings, visual analysis has a number of practical shortcomings based on the limited amount of information that can be displayed in a simple data plot, and on the limitations and general imprecision of visual perception. Mr. Higgins correctly pointed out that it is impossible to

---

[22] It is unclear to me how he could reach this conclusion in light of his claims that hours and TPH were inadequate "proxies" for costs and volumes.

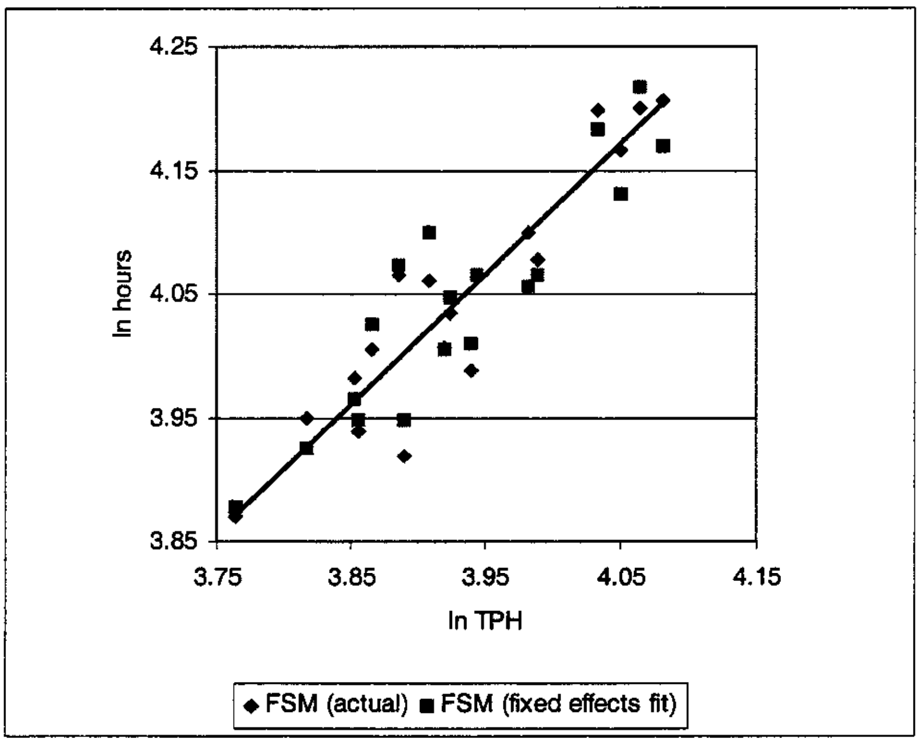
determine from the plots in Dr. Smith's Exhibit OCA–602 whether any two points represent observations of the same site in different periods, the same period at different sites, or different sites and periods (Docket No. R97–1, Tr. 33/17992–3). The same is true of the plot presented to Dr. Bradley. Plotting cross-section data, or time-series data for specific facilities, solves this problem but makes the effects of other relevant variables no more visible. Mr. Higgins raised another excellent point in stating that visually fitting a line or curve to a plot is not an adequate substitute for numerical analysis and formal specification tests. While the data in the plot presented to Dr. Bradley may appear to fall along a simple regression line, one would decisively reject the statistical hypothesis that such a line is the "true" relationship.[23] This is just equivalent to saying that variables other than TPH are relevant for explaining workhours.

Dr. Smith's efforts to classify plots for individual sites as consistent with a pooled model, fixed-effects model, or a "blob" face similar limitations. The eye can discern, albeit imprecisely, the fit (or lack thereof) of a line to a plot of data. However, the eye cannot readily ascertain how the fit of nonlinear functions or functions of several variables—the translog labor demand functions that I recommend (and Dr. Bradley also used) are both—would appear in hours-versus-TPH graphs. As a result, Dr. Smith had no way to determine whether

[23] See also Section VII.B, below.

more complicated models could fit the data better than a straight line.[24] It is easy to find cases where the data appear at first to be consistent with the "pooled" model by Dr. Smith's criteria, but actually the fixed-effects model fits the data better. Indeed, the fixed-effects model can achieve the superior fit despite the handicap that its regression coefficients (other than the intercept) are the same for every site. I show such a case in Figure 1, in which I plot FSM hours and TPF

Figure 1. Actual and fitted FSM hours and TPF, IDNUM = 3



(in natural logs, and transformed for the autocorrelation adjustment) for one site (IDNUM = 3). I also plotted the fitted hours from the fixed-effects model against

---

[24] The problem is potentially most severe for plots Dr. Smith characterized as "blobs." Seemingly random data in an hours-versus-TPH plot could simply be the two-dimensional appearance of more complicated functions of all the relevant variables.

TPF for the same observations, and the simple regression line fitted to the plotted (actual) data. The fixed-effects model provides a better fit for observations where the fixed-effects fitted value (plotted with squares) is closer than the simple regression line to the actual value (diamonds). The simple regression provides a slope of 1.045—close to one—which could be interpreted as supporting the 100 percent variability assumption for this operation. However, the fixed-effects model actually provides a much better overall fit than the straight line based only on the site's data. The mean squared error of the fitted hours from the fixed-effects model, 0.00054, is less than half the 0.00135 mean squared error of the straight line's fit. I provide the data and calculations in the srpeadsheet Figure1.xls, in Library Reference LR–I–107.

V.B. Multivariate statistical models are the only reliable means for quantifying volume-variability factors for mail processing operations

It must be recognized that the conclusion that there are numerous factors in addition to volumes that impact mail processing costs has clear implications for the validity of certain modeling approaches. It is impossible to control for the effects of various cost causing factors without including variables in the regression models that represent those factors. Inferences made from analyses that do not take into account the control variables—analyses such as the examination of "simple, unadorned" plots of costs or workhours versus TPH—will be strictly invalid, unless one of two conditions can be shown to exist. The first condition, that the explanatory variables are strictly uncorrelated, simply does not hold. The second condition, that the additional explanatory factors are actually irrelevant, can be decisively rejected on operational, theoretical, and statistical grounds. Thus, multivariate regression modeling is the only valid basis for disentangling the relationships among the various cost-causing factors, and developing testable inferences about the degree of volume-variability of mail processing costs.

The Commission cited Dr. Neels's statement that "common sense" suggests that mail processing labor costs are 100 percent volume-variable, as well as descriptions of mail processing activities (in effect, a common sense analysis) formerly used by the Postal Service to support the 100 percent variability assumption (PRC Op., Docket No. R97–1, Vol. 1, at 68–69). I believe it is necessary to view such "common sense" with a considerable degree of skepticism. While common sense can play an important role in ruling out the flatly impossible—the "laugh test"—it must first be informed as to the location of the dividing line between the impossible and the merely counterintuitive. In fact, the 100 percent variability assumption is not self-evidently true, and less-than-100 percent variabilities (equivalently, economies of density) are not even particularly counterintuitive. Consider the following:

- Economies of density are possible, according to economic theory
- Economies of density have been shown to exist in published empirical cost studies of other network industries
- Volume-variability factors less than 100 percent have been shown to exist in Postal Service cost components other than mail processing, according to volume-variability methods accepted by the Commission

Mr. Degen's testimony shows that qualitative factors that are associated with less than 100 percent volume-variability are widespread in mail processing operations. However, he is appropriately circumspect in stating that the qualitative analysis cannot quantify the degree of volume-variability (USPS-T-16 at 4). Indeed, the traditional analysis supporting the 100 percent variability assumption only reaches its quantitative conclusion by way of a simplistic model of mail processing activities (Id. at 6).

V.C. Use of the translog functional form for the mail processing labor demand models is appropriate

For this study, I chose to continue Dr. Bradley's use of the translog functional form for the mail processing labor demand models. The translog has general applicability because it provides a second order approximation to a function of arbitrary form. This allows me to place as few mathematical restrictions as possible on the functional form of the underlying cost and production functions.[25] It also permits a degree of agnosticism on the question of whether the Postal Service actually minimizes costs. As I stated in Section III.A, above, if the Postal Service were not a strict cost minimizer, I would expect the same general factors—volumes, network, wages, capital, etc.—to determine labor demand, but the effects of those factors would tend to differ from the cost minimizing case. In either case, the use of a flexible functional form is justified.

---

[25] To make a technical point, by specifying translog labor demand functions, I do not presuppose a translog cost function, which would generally imply additional restrictions on the derived factor demand functions.

Another important feature of the translog labor demand function is that it does not restrict the output elasticities (volume-variability factors) to be the same for every site or every observation, even when the slope coefficients are pooled. In contrast, if I were to have used a simpler specification such as the log-linear Cobb-Douglas functional form, then pooling the slope coefficients would restrict the variabilities to be the same for all sites. The output elasticities derived from the translog labor demand function are a linear combination of parameters and explanatory variables, and thus can vary with the level of piece handlings and other factors. (I discuss issues related to aggregating these results in Section V.F, below.) The estimated regression coefficients themselves generally do not have a simple, fixed economic interpretation.[26] Rather, the estimates must be regarded in terms of quantities such as elasticities that have an economic interpretation. Furthermore, since I do not impose any bounds on the parameter estimates from the translog functions, the elasticities may take on any value, in principle. Nothing I have done would preclude a 100 percent (or greater)

[26] The apparent exceptions prove the rule here. For example, Dr. Bradley's mean centering of the data allowed him to interpret the coefficients on the natural log of TPH as the values of the elasticity functions evaluated at the sample mean of the explanatory variables, but not as the elasticity applicable to every site or observation taken individually.

volume-variability result if that were consistent with the true structure of costs in the sorting operations.[27]

V.D. The use of the fixed-effects model offers major advantages over cross-section or pooled regressions for the mail processing models

The mail processing labor demand models must include a large number of explanatory variables in order to capture the effects of the key factors that determine mail processing labor usage. Some of the explanatory variables are correlated, so to reliably disentangle the cost effects of changes in piece handlings from other factors, it is desirable to have as much variation as possible in the data.[28] Some factors are difficult to quantify or simply unobservable, so it is necessary to employ methods that can control for them. These considerations weigh strongly in favor of the use of panel data, which offers both cross-section and time-series variation in the data, and the fixed-effects model, which can control for the effects of unobserved site-specific "fixed" factors.

---

[27] Because of the relatively large number of explanatory variables and their complicated interactions, manipulating the models to assure any given volume-variability result (as insinuated by ABA, et. al., in an interrogatory to Dr. Bradley; see Docket No. R97–1, Tr. 11/5411) would require an extraordinarily complicated set of restrictions on the parameters of the labor demand functions that are, of course, totally absent from my models and Dr. Bradley's models.

[28] It is important to note that there is no fundamental statistical problem created by including mutually correlated explanatory variables in a regression model, as long as the variables are not perfectly correlated. As Goldberger notes, multicollinearity does not alter the fundamental properties of properly specified regression estimators such as unbiasedness, or cause the estimated standard errors to be incorrect; it can even increase the precision of inferences about linear combinations of parameters. See A. Goldberger, *A Course in Econometrics*, Harvard University Press 1991, Chapter 23.

The main problem with estimation approaches such as the "pooled" or cross-section models is the difficulty in capturing the effects of the relevant explanatory variables and therefore avoiding omitted variables bias, rather than inherent inapplicability, as Dr. Bradley observed (Docket No. R97–1, Tr. 33/17907–17909). To obtain unbiased estimates from the pooled or cross-section model, it is necessary to explicitly include all explanatory variables in the regression specification, since those models lack the site-specific intercepts of the fixed-effects model, and thus cannot control for unobserved cost-causing characteristics of the sites. If some of the site-specific characteristics are not merely unobserved, but actually unobservable, the difficulty in obtaining unbiased estimates from the pooled or cross-section models becomes impossibility. Indeed, the popularity of panel data in applied productivity analysis derives substantially from the unobservability of such important factors as the quality of management.

In addition to the problem of omitted variables bias, the cross-section approach raises a number of problems caused (or at least exacerbated) by the reduction in sample size relative to a panel data approach. The number of available observations for a cross-section regression cannot exceed the number of sites—this leads to an absolute limit of about 300 observations for widely-installed Function 1 MODS operations such as manual letters. Some operations are much less widely installed, such as SPBS; any cross-section analysis of BMC operations would be greatly limited by the existence of only 21 BMCs. Indeed, for BMC operations, it would be impossible to estimate an adequately specified flexible labor demand function from pure cross-section data. The translog labor demand model with wage, capital, network, and lagged output variables has more parameters to estimate (e.g., 32 in my MODS parcel and SPBS models) than there are BMCs.[29]

Another problem with cross-section methods for mail processing is that they ignore the time series variation in the data. The time series variation in the data has two important functions. The additional variation mitigates the effects of near-multicollinearity among the explanatory variables, and thus helps reduce the sampling variation in the elasticity estimates. Cross-section methods would therefore be expected to produce estimates subject to greater sampling variation than methods that take the time series variation in the data into account. Second, and perhaps more importantly, it provides a great deal of information on the relationship between workhours, volumes, and other explanatory factors away from the average levels (as in the "between" model) or the specific levels prevailing in a particular time period. By ignoring this information, estimates based solely on cross-section information potentially have less predictive power than models incorporating all available information.

The use of panel estimators also can mitigate the effect of potential problems caused by measurement errors in the explanatory variables. The "between" model is a cross-section regression on the firm means of the data. In Docket No. R97–1, Dr. Neels claimed that errors-in-variables are less of a

[29] The same problem would exist for other functional forms, such as the generalized linear or generalized Leontief, offering a second-order approximation to an unknown function of several variables.

problem for the between model since the firm means of the data contain an averaged error (Docket No. R97–1, Tr. 28/15629). Dr. Neels's claim is only partially correct. Replacing the data with the firm means potentially reduces the variance of nonsystematic (random) errors, but it does nothing at all about systematic errors (biases) that may be present in the data. The "fixed-effects" or "within" model, in contrast, can eliminate the effects of certain systematic errors in the data. That is, since a systematic error in the data will also appear in the mean of the data, the systematic errors will tend to cancel out when the data are expressed as deviations from the individual (site) means.[30]

The fixed-effects model also can account for potential systematic errors in the workhours data that would result in omitted-variables bias for other estimators. If clocking errors were purely random, so that recorded workhours were correct on average, the only effect on the estimates would be a loss of estimator efficiency relative to the ideal case in which the hours could be observed without error.[31] However, if the reported workhours data were to reflect systematic clocking errors, such that workhours were systematically overstated or understated for certain operations and/or sites, the data would contain a non-volume "fixed effect" related to the degree of over- or understatement. This

[30] Recall that the computationally efficient mechanism for estimating the slope coefficients with the fixed-effects model is to estimate the model using deviations of the data from individual (site) means.

[31] In this case, the clocking error would become a part of the overall regression error term.

would result in omitted-variables bias in the cross-section and pooled estimators, but not the fixed-effects estimator.

V.E. No regression method inherently embodies any given "length of run"

There is no general point of economic or econometric theory implying that any given regression technique—pooled, cross-section, time series, fixed- or random-effects, etc.—yields inherently "shorter run" or "longer run" results than another. Econometrics texts are devoid of generalities that prescribe a particular data frequency or extent of time aggregation of the data for a given type of econometric analysis.[32]

Some might justify a preference for cross-section analysis on the idea that differences between cross-sectional units (i.e., sites) reflect "long-run equilibrium" differences. There are, in fact, two significant assumptions underlying such a view, neither of which is applicable to Postal Service operations. First, it assumes that mail processing operations are actually observed "in equilibrium"—it is doubtful that they are, given the dynamic staffing adjustment processes. If sites could somehow be observed in their long-run equilibrium states, it would still not militate in favor of cross-section analysis: time-series comparisons would be no less indicative of long-run cost variations than cross-section comparisons. Second, and more importantly, it assumes that even if the operations could be

[32] It is also not necessary that the sample period and the "rate cycle" coincide. In Docket No. R71–1 the Chief Examiner chided the Postal Service for "confus[ing] the test period with the period for determining variability... they measure different things and need not coincide" (Docket No. R71–1, Chief Examiner's Initial Decision, at 23).

observed in long-run equilibrium, there would be no non-volume differences between sites. In fact, it is evident that there are highly persistent non-volume differences between plants for which controls will be needed in any scenario relevant for Postal Service costing. It is not necessarily impossible to contemplate a "long run" in which the present diversity of big- and small-city facilities will be replaced by homogeneous operations, but it is clear that that such a "long run" is many rate cycles distant.

To forge ahead and estimate a long-run cost function from cross-section data when the data are not observed in long-run equilibrium results, as Friedlaender and Spady point out, in biased estimates of the relevant economic quantities (see A. Friedlaender and R. Spady, *Freight Transport Regulation*, MIT Press 1981, p. 17).

V.F. The "arithmetic mean" method is an appropriate technique for aggregating the elasticity estimates; using alternative aggregation methods from past proceedings does not materially impact the overall results

Mail processing operations differ widely in their output levels and other cost causing characteristics. The degree to which costs are responsive to changes in mail volumes may, therefore, vary from activity to activity[33] and also from site to site. I take this into account by estimating separate labor demand functions for each sorting activity, and by using a flexible functional form that allows the elasticities to vary with the characteristics of each observation. The

[33] The Data Quality Study also expresses this opinion (Data Quality Study, Summary Report, at 76).

elasticities are able to vary over sites and time periods even though slope coefficients of the translog function are "pooled." The pooling restrictions on the regression slopes in the labor demand models do not imply similar restrictions on the elasticities, which are the economic quantities of interest.[34] For use in the CRA, it is necessary to determine a nationally representative, or aggregate, value of the elasticities.

Below, I use the term "aggregate" to refer generically to any method by which the elasticity formulas are evaluated at representative values of the variables, or by which individual elasticities generated using the formulas are combined or averaged into a national figure. My usage of the term differs from Dr. Bradley's in Docket No. R90–1, where Dr. Bradley collectively termed the "average-of-the-variabilities" methods the "disaggregated" approach (Docket No. R90–1, Tr. 41/22061).

Typically, the aggregate values of econometrically estimated elasticities employed in the CRA have been computed by evaluating the elasticity functions at the sample mean values of the relevant explanatory variables (I refer to this below as the "arithmetic mean method"). As a means of obtaining representative values for the elasticities, the arithmetic mean method has clear intuitive appeal, as the arithmetic mean is a common and simple way to determine system

---

[34] This point was also made by Dr. Bradley in Docket No. R97–1 (see Docket No. R97–1, Tr. 28/16073).

average values for the explanatory variables.[35] For his mail processing study in Docket No. R97–1, Dr. Bradley used the arithmetic mean method to evaluate the elasticities. He implemented the approach by "mean centering" his data prior to estimating his regression models, which has the effect that the estimated regression coefficient on the natural log of TPH is equal to the desired aggregate elasticity of labor demand with respect to the operation's output.

In reviewing Dr. Bradley's study, the Commission expressed a concern that the aggregate elasticities computed by Dr. Bradley might be inapplicable to particular facilities (PRC Op., R97–1, Vol. 1, at 91). The Commission's concern is correct in the sense that the aggregate elasticity is not necessarily the best predictor of an individual site's cost response to a volume change, much as the national unemployment rate (or any other national aggregate economic statistic) does not necessarily reflect conditions in a specific locality. However, this apparent "problem" is a deliberate feature of the analysis, because the aggregate elasticities are meant to represent a systemwide response. This is consistent with the goal of the costing exercise, which is to determine composite or nationally representative cost responses to representative volume-related workload changes. As Mr. Degen explains, it would be inappropriate to assume that national (RPW) volume changes on the margin would be concentrated in a sufficiently small number of origin-destination pairs to make the degree of variability at one or a few facilities unusually important. Rather, a national

[35] Additional merits of the arithmetic mean method are discussed at some length in Dr. Bradley's rebuttal testimony from Docket No. R90–1 (Docket No. R90–1, Tr. 41/22052–22061).

volume change will tend to affect workload at every site (see USPS–T–16, at 15–17). As a result, the elasticities for the individual sites are of interest primarily for their contribution to the systemwide cost response.

Nevertheless, the elasticities for individual sites and/or observations have a useful diagnostic function. A model that produces reasonable results when the elasticities are evaluated at the mean may well produce unreasonable results when evaluated at more extreme (but still plausible) values in the data set. Such problems may not be evident from standard goodness-of-fit statistics. Dr. Bradley's mean centering method is a convenient way to obtain the aggregate elasticities, but it interferes with the task of computing estimated elasticities for individual sites and/or observations.

The arithmetic mean method is not the only theoretically valid way to compute aggregate elasticities. In Docket No. R90–1, the Commission considered a variety of elasticity aggregation methods as part of its review of the cost analyses for the city carrier street components. In that proceeding, Dr. Bradley advocated the arithmetic mean method, which had been accepted by the Commission in Docket No. R87–1 and employed in other cost components. Several intervenors countered with "average-of-the-variabilities" methods, in which values of the elasticities are generated for each observation and the results averaged with or without weights. The Commission correctly concluded that the arithmetic mean method was applicable, but that there was some substance to the intervenor alternatives, so methods other than the arithmetic mean are justifiable under some circumstances (PRC Op., Docket No. R90–1, Vol. 1, at III–16).

To facilitate examination of the distributions of individual elasticities, as well as comparisons of the results from alternative aggregation methods to the arithmetic mean, I chose not to use Dr. Bradley's mean centering approach. Rather, I explicitly derived the elasticity formulas from the translog factor demand functions I estimated, and explicitly calculated the elasticities using those formulas.

A special property of the translog factor demand function is that the output elasticity is a linear combination of the natural log of the explanatory variables.[36] Thus, it is straightforward to compute variances (conditional on the explanatory variables) for the aggregate elasticity estimates using the covariance matrix of the model parameters and standard formulas for the variance of a linear combination of random variables. I provide estimated standard errors along with the elasticities in the results I report in Section VII.A, below.

I considered three elasticity aggregation methods. These are the arithmetic mean method, which I recommend using for the Postal Service's Base Year 1998 mail processing costs, a variation on the arithmetic mean method using the geometric mean in place of the arithmetic mean, and a weighted geometric mean method (using workhours as weights). These methods

---

[36] Formally, for a translog function with N explanatory variables $y = \alpha + \sum_{n=1}^{N} \alpha_n \ln x_n + \sum_{m=1}^{N} \sum_{n=1}^{N} \alpha_{mn} \ln x_m \ln x_n$, the elasticity of y with respect to the jth explanatory variable is $\varepsilon_j = \partial \ln y / \partial \ln x_j = \alpha_j + \sum_{n=1}^{N} \alpha_{jn} \ln x_n$.

encompass the alternatives proposed in Docket No. R90–1. Given the form of the mail processing elasticity equations, it can be shown that the geometric mean method is algebraically equivalent to the unweighted average elasticity method proposed in Docket No. R90–1 by MOAA et al. witness Andrew; the weighted average elasticity methods proposed in Docket No. R90–1 by Advo witness Lerner and UPS witness Nelson are equivalent to variations on the weighted geometric mean method.[37] I also compared elasticities computed from the full regression samples with elasticities derived using only the FY1998 subset of observations. The mathematical details of the methods are presented in Appendix E.

The equivalence of the "average-of-the-variabilities" and geometric mean methods means that the differences in the methods boil down to the differences in the arithmetic and geometric means as measures of the characteristics of the representative facility. The geometric mean typically is less sensitive to extreme values of the data than the arithmetic mean, which may make it a more suitable measure of central tendency for skewed or otherwise long-tailed data. However, since large facilities—whose data are in the upper tails of the distributions of certain explanatory variables, particularly TPF and possible deliveries—will tend to represent a large share of costs, it may be undesirable to implicitly de-emphasize them using an unweighted geometric mean method. Consequently, the hours-weighted geometric mean may be preferable among the geometric

[37] See Docket No. R90–1, Tr. N/22061–22063. The methods advocated by witnesses Lerner and Nelson differ in their use of, respectively, predicted and actual costs as weights. I did not compute the elasticities using Lerner's method.

mean methods. The hours-weighted geometric mean method also has the theoretical feature that it is synonymous with aggregating marginal cost using TPF weights. If it were believed that, other things equal, the distribution of marginal TPF (mail processing volumes) across facilities resembles the distribution of existing TPF, then the weighted geometric mean method would be appropriate.[38] However, uncertainty as to the actual geographical pattern of marginal mail processing volumes makes it less clear that the weighted geometric mean method (or any other weighting approach) is superior. Consequently, I chose to retain the arithmetic mean method to compute the aggregate elasticities.

Another issue is the appropriate way, if any, to use data from previous years to evaluate the elasticities for the 1998 Base Year. Note that this is a separate issue from the issue of whether it is appropriate to use previous years' observations to estimate the labor demand functions. While the FY1998 observations may be, in principle, the best measures of conditions prevailing in the Base Year, there are some complications that may weigh in favor of using additional data. For example, the sizes of the 1998 subsamples are considerably smaller than the total sample sizes for all operations; small sample instability of the site means would be a concern. The Base Year might not be representative of conditions likely to prevail over the "rate cycle," for instance if the Base Year happens to be a business cycle peak or trough. Finally, there are statistical

[38] Major geographical shifts of mail processing volumes will tend to be driven by corresponding shifts in the population served by the Postal Service, rather than by the volumes in themselves.

advantages to evaluating the elasticities at the overall sample means of the variables (see PRC Op., Docket No. R90–1, Vol. 1, at III–11 to III–12). As with the arithmetic-versus-geometric mean decision, I do not find the case for using only the FY1998 observations to be sufficiently compelling to recommend the elasticities based solely on the 1998 observations.

The composite volume-variable cost percentages from the six aggregation methods for the cost pools with econometrically estimated elasticities, presented in Table 2, fall in the remarkably narrow range of 74.7 percent to 76.0 percent. No method provides uniformly higher measured elasticities for every cost pool (see Appendix D). The elasticities based on the 1998 subsets of observations are, on balance, slightly lower than the elasticities based on the full regression samples. I conclude that the choice of aggregation method does not greatly impact the overall volume-variable cost percentage for the cost pools with econometrically estimated labor demand functions.

Table 2. Comparison of Composite Volume-Variable Cost Percentages for Selected Aggregation Methods

Observation Set	Aggregation Method	Composite "Variability"
Full regression sample	Arithmetic mean (BY 1998 USPS method)	76.0%
	Geometric mean	75.1%
	Weighted geo. Mean	74.9%
FY1998 subset of regression sample	Arithmetic mean	75.6%
	Geometric mean	74.8%
	Weighted geo. Mean	74.7%

Source: Appendix D, Tables D–1 and D–2.

V.G. Eliminating grossly erroneous observations is acceptable statistical practice

A fundamental, if easy to overlook, fact is that it is not necessary to use every observation that is (in principle) available to the researcher to draw valid inferences from regression analysis or other statistical procedures. Using a subset of the potentially admissible observations generally has adverse consequences for the efficiency of estimation, relative to the case in which all observations are employed, but does not generally result in bias (or inconsistency).[39] The importance of this fact is easiest to see in the alternative—if it were not so, it would be impossible to conduct any admissible statistical analysis without perfectly complete and correct data. In practice, no data collection system can be presumed to be perfect, and since many common statistical procedures may break down in the presence of grossly erroneous data, the normal situation is that such data must be identified and then eliminated or corrected before the analysis may proceed.

I do not intend to suggest that all data "outliers" can be discarded with impunity. Indeed, the prescriptions of the statistics literature for outlying but correct observations differ considerably from those applicable to grossly erroneous observations. The presence of outlying but correct observations may signal the need to specify a more general (or otherwise different) model that can better explain the outlying observations. However, when the data are erroneous,

[39] The result can be proved in a variety of ways. For example, it is a straightforward byproduct of the derivation of the F-test for the equality of two regression populations. See P. Schmidt, *Econometrics*, Marcel Dekker, Inc., New York, 1976, pp. 29–30.

removing them is usually warranted, to reduce the likelihood they will induce serious errors (in either direction) in the estimated relationships. One text on robust statistical methods puts it rather bluntly: "Any way of treating [i.e., rejecting] outliers which is not totally inappropriate, prevents the worst" (F. Hampel, et. al., *Robust Statistics*, John Wiley & Sons, New York, 1986, p. 70). By "totally inappropriate," they mean methods that identify and reject outliers on the basis of non-robust statistics, such as the residuals from least squares regressions. I do not use, nor did Dr. Bradley use, any such methods to determine regression samples. Even Cook and Weisberg—quoted by Dr. Neels as saying, "[i]nfluential cases... can provide more important information than most other cases" (Docket No. R97–1, Tr. 28/15613)—clarify (in the paragraph immediately following the passage excerpted by Dr. Neels) that there is no reason to retain grossly erroneous observations:

> If the influential cases correspond to gross measurement errors, recording or keypunching errors, or inappropriate experimental conditions, then they should be deleted or, if possible, corrected (R. Cook and S. Weisberg, *Residuals and Influence in Regression*, Chapman and Hall, New York, 1982, p. 104).

While correcting the data may be preferable, it is infeasible in the case of the MODS data, since it is not possible to re-count the pieces. Even corrections were feasible, they would tend to be prohibitively expensive.[40] Thus, I conclude that removing grossly erroneous data from the regression samples is acceptable practice.

---

[40] Note that it may be feasible, and far more economical, to apply statistical procedures to mitigate measurement errors.

The practical relevance of the Cook and Weisberg statement about the potential importance of "influential cases" is further limited in the context of Postal Service operations. For some purposes, such as investigating chemical compounds for antibiotic properties, the outliers—the presumptive minority of compounds that "work"—may actually be the only observations of interest to the researcher (see Hampel, et al., p. 57). Of course, in such a case, the outliers are probably not erroneous observations. Extremely unusual observations of the Postal Service's mail sorting activities are far less likely to be correct, since the organization of the activities is reasonably well known. If an observation of a Postal Service sorting operation suggests that the operation is expanding greatly the boundaries of human achievement, or is a black hole of idle labor, I believe it is safest by far to conclude that the observation is a gross error. Furthermore, it does not suffice to know the how the labor demand in an activity responds to volumes at only one facility or a handful of facilities, however remarkable those facilities may be. Since variations in mail volumes will tend to cause variations in labor demand throughout the mail processing system, the "mundane" facilities will, therefore, contribute some—conceivably the largest—portion of the system-wide labor cost response to a national volume change.[41]

---

[41] One situation Dr. Neels may have had in mind is a situation in which a few sites represent the leading edge of a planned technology deployment. Of course, if further deployment is anticipated to lead to materially reduced costs in the test year relative to the base year, it should be reflected in a cost reduction in the rollforward model. Nonetheless, the other sites' observations would appropriately reflect the technology in place as of the base year.

V.H. The potential consequences of erroneous data for the labor demand models vary according to the type of error

The discussion in the previous section should not be interpreted as suggesting a blanket need to eliminate all erroneous observations from the model. In Docket No. R97–1, the Commission stated that "removing even erroneous data from a sample without investigating for cause is not representative of the best econometric practice" (PRC Op., R97–1, Vol. 1, at 84). In some respects, I agree—the real issue is not the presence of errors *per se*, but rather their materiality for the estimation process. Some types of data errors are inconsequential because they would have no bias or inconsistency effect on regression estimates whatsoever, while other types of errors are potentially problematic but may be handled using appropriate modeling techniques. I will discuss potential measurement errors in both workhours and piece handlings separately, and in each case distinguish the potential effects of random (nonsystematic) errors from those of biases (systematic errors).

V.H.1. Errors in workhours

Neither type of potential measurement error in workhours figured in the intervenor criticism of Dr. Bradley's study. I believe this is largely due to the statistical fact that random error in the dependent variable of a regression cannot be distinguished from the usual regression disturbance, and thus does not lead to biased or inconsistent estimates of the regression coefficients (see P. Schmidt, *Econometrics*, Marcel Dekker, 1976, p. 106).

Certain types of systematic errors in workhours can, however, cause the pooled and cross-section models to produce biased and inconsistent coefficient estimates. The problem arises if there are systematic errors in workhours that vary in degree by site. If systematic mis-clocking occurs, I do not believe there is any reason why all sites should make the same errors. The underlying causes of the errors would likely be idiosyncratic to the sites. For instance, if the errors were deliberate, different sites presumably would face different incentives to shift their workhours among the operations. In any event, the measured workhours for a given site would differ from actual workhours by a site-specific factor.[42] Since neither the pooled model nor the cross-section model controls for such site-specific effects, estimates from those models would be subject to misspecification bias. In contrast, the fixed-effects model is robust to this type of bias in workhours since the systematic clocking errors would simply be absorbed in the site-specific intercepts.

V.H.2. Errors in piece handlings

The effects of errors in the piece handling data on the results also differ according to whether the errors in the data are random or systematic. Since piece handlings are an explanatory variable, random errors in the data are not as innocuous as they are in the case of random errors in workhours. Dr. Neels criticized Dr. Bradley because "measurement error in an independent variable

[42] Formally, $HRS_{it}^{observed} = \eta_i \cdot HRS_{it}^{actual}$, or in logs $\ln HRS_{it}^{observed} = \eta_i^* + \ln HRS_{it}^{actual}$, where $\eta_i^* = \ln \eta_i$.

causes downward bias in coefficient estimates" (Docket No. R97–1, Tr. 28/15604). That Dr. Neels failed to distinguish between random and systematic measurement error is evident from Dr. Neels's subsequent discussion of the problem. The downward bias or "attenuation" effect Dr. Neels described is, more precisely, a result of random measurement error, and his prescription of the between model (cross-section regression on site means) to mitigate the effect of possible measurement error is useful only for random measurement errors. Another consequence of Dr. Neels's conflation of random error with all measurement error is that he failed to consider the possibility that errors in piece handlings could have primarily taken the form of systematic error, rather than random error. This is an important case to consider, since the theory behind the attenuation result indicates that (relatively) small random errors cause small biases in the regression coefficient estimates. Furthermore, the between model prescribed by Dr. Neels is not robust to the presence of systematic errors in the explanatory variables.

The most severe potential errors in the manual[43] piece handling data are likely to be systematic, rather than random, errors. The main source of error in manual piece handlings is the use of weight conversions and "downflow densities" for their measurement. The conversion error has two statistically distinct components—a random error inherent to the conversion process, and a potential systematic error (bias) resulting from the application of outdated or

---

[43] Recall that unlike mechanized and automated operations MODS TPH and TPF for manual operations are obtained from weight conversions. See Postal Service Library Reference LR–H–147 from Docket No. R97–1.

otherwise incorrect conversion factors. The Inspection Service's report on the Postal Service's volume measurement systems (Docket No. R97–1, USPS–LR–H–220) focused exclusively on sources of systematic error, or bias, in FHP measurement.

The main potential problems with the national weight conversion factors and the downflows—respectively, local differences in weight per piece from national averages,[44] and the accuracy and currency of the locally generated densities—would tend to vary in severity from site to site. The prospect of site-specific biases clouds cross-sectional comparisons of piece handlings, since any measured difference (or lack thereof) would be the result of a combination of the difference (if any) in actual handlings and the differential bias. A cross-section regression on means such as the between model offers no relief, since if the observations of piece handlings are biased, the average piece handlings will also be biased. In contrast, the "within" transformation (representing the data as deviations from site means) used to implement the fixed-effects model, automatically sweeps out site-specific biases from the data.[45] As a result, the

---

[44] Since the weight conversions are applied by "source/types" that subdivide the letter and flat mailstreams, differences in local weight per piece from the national average will only cause a problem if there are differences below the source/type level.

[45] Formally, given an observed regressor $x_{it}^{*} = v_i + x_{it}$, where x_{it} is the actual value of the data and v_i the site-specific bias, the site means $\bar{x}_i^{*} = v_i + \bar{x}_i$ used in the between model are biased, but the deviations $x_{it}^{*} - \bar{x}_i^{*} = v_i + x_{it} - (v_i + \bar{x}_i) = x_{it} - \bar{x}_i$ used in the fixed-effects model are not.

fixed-effects model will be more robust to the presence of biased data than the between model.

While random measurement error in explanatory variables can lead to downward bias in regression coefficient estimates, the evidence on the record in Docket No. R97–1 indicates that the random components of measurement errors in piece handlings generally have small variances and correspondingly small effects on the estimates (see Docket No. R97–1, USPS–T–14, at 83–84; Tr. 33/17897–17900, 18009–18012, 18014–18019). Dr. Neels erroneously attempted to discredit the "errors-in-variables" findings as embodying a "mathematically impossible" result of negative estimated measurement error variances—a result which Mr. Higgins and Dr. Bradley correctly identified as entirely possible in finite samples (Docket No. R97–1, Tr. 33/17898–17900, 18016–18017). In fact, the purported anomaly would be most likely to occur in situations where the bias due to random measurement error is inconsequential. Because of sampling variability, an errors-in-variables point estimate can be lower than the corresponding fixed-effects point estimate even though the fixed-effects result would tend to be lower on average because of the bias. This result would be highly improbable if the actual bias were large relative to the sampling error variance.

VI. Data

VI.A. Data Requirements for Study

The analysis in Sections III and IV, above, indicates that MODS data alone are not sufficient for estimation of labor demand functions for mail processing operations. In addition to MODS data on workhours, or real labor input, and piece handlings, or mail processing volumes, I require data to quantify characteristics of the sites' local service territory and the economic variables of wages and capital input. I briefly describe the MODS data in Section VI.B and the data sources other than MODS in Section VI.C, below.

VI.B. MODS Data

The MODS data I employ are similar to the data employed by Dr. Bradley in Docket No. R97–1. I aggregate the MODS workhour and piece handling data from the three-digit operation code level to the mail processing cost pool groups employed for cost distribution purposes in both the Commission and the Postal Service methods. Based on Ms. Kingsley and Mr. Degen's descriptions, the mail processing cost pools established in Docket No. R97-1 continue to reflect the important technological distinctions among sorting operations and are generally appropriate for volume-variability estimation. However, I also aggregated the SPBS Priority and non-Priority cost pools into a combined SPBS pool, since the divergent variability results without a clear operational basis suggested that the more detailed cost pools may have been too finely drawn for variability estimation.[46] As I describe in Section IV.D, above, in the automated and mechanized sorting operations (BCS, OCR, FSM, LSM, and SPBS), Total Pieces Fed (TPF) is a better measure of piece handlings than Total Pieces Handled (TPH), since the former includes rejected pieces in the total output. I collected both TPH and TPF data for the automated and mechanized sorting operations.

VI.C. Other Postal Service Data

In order to build a data set with sufficient information to estimate the mail processing labor demand models described in Sections IV and V, I employed data from several Postal Service data systems in addition to MODS. The systems include the National Workhour Reporting System (NWRS), Address Information System (AIS), Address List Management System (ALMS), Facility Management System (FMS), Installation Master File (IMF), National Consolidated Trial Balance (NCTB), Personal Property Asset Master (PPAM), and Rural Route Master (RRMAS).

VI.C.1 Delivery Network Data—AIS, ALMS, RRMAS

AIS records the number of possible deliveries by delivery type (e.g., centralized, curbline, NDCBU), route, and Finance number. AIS data are collected by carriers, who record the number of deliveries at each stop on the route. The detailed data are entered into the AIS system, and AIS software

---

[46] However, mail mix differences between the detailed SPBS cost pools favor retaining the distinction for distribution key formation.

calculates the total deliveries for the route.[47] Unlike most of the other data I use, the AIS data are collected by month rather than by accounting period.[48] The process by which the monthly data are mapped to postal quarters is described in LR–I–107.

ALMS contains information for each post office, station, and branch by Finance number. This information includes a contact name and telephone number, the address, CAG, facility type (e.g. station, branch, post office), and ZIP code. A station is a unit of a main post office located within the corporate limits of the city or town while a branch is outside the corporate limits. It also distinguishes contract facilities from non-contract facilities (contract facilities do not have Postal Service employees). Four variables are created from ALMS: number of large post offices, number of small post offices, number of stations and branches, and number of 5-digit ZIP Codes in each REGPO. A large post office is defined as a Class 1 or Class 2 post office. A small post office is defined as a Class 3 or Class 4 post office.

RRMAS contains information on rural route deliveries by route and finance number. RRMAS is used to create the total number of rural deliveries by REGPO. This information is also available in AIS but the data in RRMAS is believed to be more accurate for rural deliveries. Rural boxes can be double counted in AIS if the route involves a stop at an intermediate office or deliveries

---

[47] Prior to AP6 of FY1994, total deliveries by delivery type and route were entered into AIS, rather than the detailed data now collected by the carriers.

[48] The data for FY1993 and part of FY1994 had AP frequency.

to boxes on another route. Additionally, RRMAS data are used in the rural route evaluations.

For all delivery network variables, the data are rolled up to 3-digit ZIP Code. The 3-digit ZIP Code data are then mapped to REGPOs using a destinating mail processing scheme. The destinating mail processing scheme is based on a map developed by the Postal Service to indicate which facility processes destinating First-Class Mail for each 3-digit ZIP Code. It is then straightforward to map Finance numbers to REGPOs. The map is updated when obvious changes to the scheme occur (e.g. plant closings or openings). Not all 3-digit ZIP Codes get mapped to a facility. In these cases, the First-Class Mail for the 3-digit ZIP Code is assumed to be processed locally or by a facility that does not report MODS. This is why several REGPOs have no delivery data mapped to them.

VI.C.2 Wage Data—NWRS

I used NWRS to obtain wage rates by site as close to the operation level as possible. MODS provides data on workhours, but not compensation amounts, by three-digit operation number. NWRS provides data on workhours and compensation amounts in dollars by Labor Distribution Code (LDC) and Finance number. The implicit wage in NWRS is the ratio of compensation dollars to workhours. Each three-digit MODS operation number is mapped to an LDC. A collection of MODS operation numbers, comprising one or more mail processing cost pools, is therefore associated with each LDC (see USPS–T–17 for details).

Since many LDCs encompass operations from several distinct mail processing streams—e.g., LDC 14 consists of manual sorting operations in the letter, flat, parcel, and Priority Mail processing streams—it is not appropriate to use LDCs as the units of production for the labor demand analysis. However, most of the important differences in compensation at the cost pool level (due to skill levels, pay grades, etc.) are related to the type of technology (manual, mechanized, or automated) and therefore are present in the LDC-level data. Thus, the LDC wage is a reasonable estimate of the cost pool-specific wage.

NWRS compensation totals tie to the salary and benefits accounts in the NCTB. As with other Postal Service accounting systems, erroneous data in NWRS sometimes arise as a result of accounting adjustments. The adjustments are usually too small to materially affect the wage calculations, but occasional large accounting adjustments result in negative reported hours and/or dollars for certain observations. Unfortunately, it is not possible to isolate the accounting adjustments. As a result, I employed procedures to identify NWRS observations with negative values of hours and/or dollars and to treat those observations as missing.

VI.C.3. Accounting data—NCTB

NCTB is an accounting data system that records the Postal Service's revenues, expenses, assets, and liabilities. NCTB data are available by general ledger account, Finance number, and AP. The data are provided as Year-To-Date totals through the current AP, which may include prior period adjustments.

While most adjustments are small relative to the current period entries, occasional large adjustments result in negative current expenses net of the adjustments. NCTB is the source for materials, building occupancy, equipment rental, and transportation expenses.

VI.C.4. Capital Data—FMS, PPAM, IMF

The Facility Master System (FMS) provides quarterly rented and owned square footage for each Postal Service facility. The beginning-of-the-year owned square footage is rolled up to REGPO, which is then used to split out the quarterly national building occupancy expenses from NCTB. The FMS data include some duplicate records and "dropouts" (e.g., a record exists for a facility in FY1996 and FY1998, but not FY1997). To obtain accurate data from the system, I employ procedures to eliminate duplicate records and interpolate missing records. These procedures are described in LR–I–107.

The PPAM is a log of equipment that is currently in use. Each record on the tape is a piece of equipment. Retrofits to existing equipment are recorded as separate records. PPAM contains the Finance number, CAG, BA, Property Code Number (PCN), year of acquisition, and cost for each piece of equipment. The PPAM data have AP frequency. PPAM classifies Postal Service equipment as Customer Service Equipment (CSE), Postal Support Equipment (PSE), Automated Handling Equipment (AHE), and Mechanized Handling Equipment (MHE). Since each PPAM equipment category encompasses a variety of equipment types, there is no simple correspondence between the categories and specific mail processing cost pool. Using the year of acquisition, the value of each year's equipment is depreciated using a 1.5 declining balance rate of replacement. For CSE, PSE, AHE, and MHE the average lives are 14 years, 13 years, 18 years, and 18 years, respectively. The annual depreciation rates are then .107 for CSE, .115 for PSE, and .083 for AHE and MHE. These depreciated values are then deflated to 1972 dollars by using annual national deflators. The annual national deflators are derived from various public and private data sources, as well as USPS sources. The deflated values from 1968 to the current year are then added together to create a total value of the equipment type in 1972 dollars. The deflated values are used as shares to distribute quarterly NCTB expenses for each equipment type.

The IMF lists the Postal Service's active Finance numbers. There are approximately 32,000 Finance numbers currently. The IMF includes details about each finance number's postal address, ZIP Code, and BA code. The BA code identifies the function (e.g., mail processing, customer services) served by each Finance number. Many of the Postal Service's databases are organized by Finance number. IMF data are instrumental in cross-walking data organized by Finance number to ZIP Codes, and thus for matching databases organized by ZIP Code with databases organized by Finance number.

VI.D. Critique of Bradley's data "scrubs"

The sample selection rules or "scrubs" that Dr. Bradley applied to the MODS data set were extensively criticized by Dr. Neels and the Commission for their liberal deletion of data and the resulting possibility of sample selection bias. I concur with Dr. Neels to the extent that certain details of Dr. Bradley's procedures are difficult to justify objectively. However, as I indicated in Section II.B, Dr. Neels's own re-estimation of Dr. Bradley's models on "all usable observations" did not demonstrate a single direction of change in the results, and Mr. Higgins further showed in his testimony that the relative magnitude of the scrubs' effect was not large. Thus, I believe Dr. Ying was fundamentally correct in stating that Dr. Bradley's sample selection rules did not build any obvious bias into his results. Nonetheless, the sample selection procedures merit re-examination to determine whether alternate sample selection rules, which might be more or less restrictive than Dr. Bradley's, might better serve the purpose of identifying the most reliable data for estimating the volume-variability factors.

VI.D.1. "Threshold" scrub

The "threshold" scrub eliminated from Dr. Bradley's data sets observations for which the reported TPH did not exceed a threshold level. Dr. Bradley set thresholds of 100,000 TPH per AP for letter and flat sorting operations, and 15,000 TPH per AP for the parcel, bundle, and cancellation operations. The difference in the threshold levels was meant to accommodate the lower volume of handlings performed in the latter group of operations (Docket No. R97–1, Tr. 11/5381, 5433). Dr. Bradley interpreted the effect of the scrub as eliminating observations for which the operation might be "ramping up" to a normal level of operation (Docket No. R97–1, USPS–T–14, at 30). The justification for eliminating the "ramping up" observations was that such observations would not be representative of the "normal operating environment" expected to prevail for the activity, and could contribute to biased measurements of the "actual" marginal cost of handlings in the operation going forward.[49]

The potential restrictiveness of the threshold scrub as applied by Dr. Bradley varies by cost pool. Processing 100,000 pieces could require fewer than ten workhours on an OCR or BCS, but well over one hundred workhours at a manual case.[50] Similarly, 15,000 piece handlings would require considerably more labor time in a manual parcel distribution than a mechanized or automated cancellation operation. Keeping in mind that the operations in question are located at mail processing plants and not small post offices, it is clear that a few workhours per day—or less—would not constitute normal levels of activity for operations in most of the sorting cost pools. The possible exceptions may be the manual Priority Mail and parcel operations, which tend to operate at low volumes since much of the sorting workload is handled in operations at other types of facilities—BMCs, PMPCs, and stations/branches (in LDC 43 operations). Dr. Bradley could have fine-tuned the threshold levels to help ensure that they

---

[49] Dr. Bradley also observed that the direction of bias is potentially ambiguous. For instance, only the "cleanest" mail might be worked on new automation equipment during the ramping-up period, which could increase productivity relative to normal operations in which more borderline automation-compatible mail would be worked. See Docket No. R97–1, Tr. 11/5354-5.

[50] For instance, assuming a machine throughput of 30,000 TPF/hour and a complement of two, processing 100,000 pieces on automation would require 6-2/3 workhours, not counting setup and takedown time. Median productivity for manual letter and manual flat sorting operation is well under 1000 pieces per workhour.

did not inadvertently exclude data from small but regular operations solely because of their smallness. Still, the median TPH per AP in my data set is 528,000 for the manual Priority Mail pool and 389,000 for the manual parcel pool—both more than 25 times greater than Dr. Bradley's threshold, even for these relatively low volume operations. Therefore, I conclude that Dr. Bradley was justified in considering observations below the threshold as highly atypical of normal operating conditions for the activities, if not actually erroneous.

However, there is nothing about the threshold scrub that indicates to me that it would remove only, or even primarily, "ramping up" observations from the MODS data set. First, the thresholds are so low that I would expect that the vast majority of plants would run these operations well above the threshold level, even while "ramping up." Second, the threshold scrub will remove observations below the threshold output level regardless of whether they are actually at the start of the operation's data. Therefore, I conclude that the actual function of the threshold scrub is—with the caveats above regarding certain small operations—to remove "noise" from the data, likely resulting from stray data entry errors. Since removing "noise" from the data is a legitimate goal of a sample selection rule, I conclude that Dr. Bradley's inclusion of a threshold check was valid in principle, though not necessarily in the details of its implementation.

VI.D.2. "Continuity" scrub

Dr. Bradley applied a "continuity" check, requiring the data for each site to be part of an uninterrupted sequence of at least thirty-nine consecutive AP observations,[51] on the grounds that "[c]ontinuous data facilitate the estimation of accurate seasonal effects, secular non-volume trends, and serial correlation corrections" (Docket No. R97–1, USPS–T–14, at 31). Dr. Neels did not take issue with Dr. Bradley's justification of the continuity check per se, but contended that it was "especially arbitrary" in its application (Docket No. R97–1, Tr. 28/15615).

To the extent that the regression models employ previous periods' data (e.g., lagged volumes) as explanatory variables or are adjusted for the presence of autocorrelation in the regression error term, some continuity of the data is required for estimation. Dr. Bradley's specification of his preferred regression models—which included a single period lag of TPH and an error term allowing for First-order autocorrelation—required that every observation in the regression sample have valid data (for the relevant variables) in the previous period. This imposed a requirement that any observation included in the regression sample be part of a block of at least two continuous APs of valid data. I do not believe there is any question that a continuity requirement derived from the regression specification is valid (see Docket No. R97–1, Tr. 28/15616), so for the remainder of this section, I refer to continuity requirements that go beyond statistical necessity.

Indeed, Dr. Bradley's continuity checks exceed statistical necessity by a conspicuously large amount, and in three distinct characteristics. First,

---

[51] Dr. Bradley also applied a version of this check requiring twenty-six consecutive APs in an exercise presented in his rebuttal testimony. See Docket No. R97–1, Tr. 33/17892–17893.

Dr. Bradley required a minimum of thirty-nine APs (three Postal Fiscal Years) of consecutive valid observations, compared to the necessary minimum of two or three. Second, if a site's data consisted of two blocks of data, both comprising at least thirty-nine consecutive APs of valid data, Dr. Bradley only included the chronologically later block in his regression sample. Third, Dr. Bradley ran his continuity checks twice—both before and after the productivity check.

Dr. Bradley offered several arguments to justify the application of a restrictive continuity check. In addition to his primary justification recounted above, Dr. Bradley claimed that data from sites that report more consistently could be presumed to be of higher quality than data from sites that report data only intermittently. Also, Dr. Bradley stated that the large number of MODS data observations gave him the freedom to be more selective about the observations he admitted into the regression samples. Each of these arguments is correct in some sense. Nonetheless, none of them inexorably led Dr. Bradley to his stringent continuity procedure—he remained free, in principle, to choose a less restrictive rule.

Beyond the continuity requirements imposed by the specification of the regression models, the main sense in which the continuity check facilitates estimation is computational. After the continuity check—more specifically, the requirement that only the most recent block of data passing the continuity check for any site be admitted to the sample—each site's data is free of reporting gaps. This was important for Dr. Bradley because he implemented his regressions using matrix calculations in SAS IML. If the continuity check were relaxed to allow reporting gaps in the data for some sites, the matrix algebra required to implement the regressions would be considerably more intricate, and the corresponding IML programming much more complex. The obvious solution is to substitute for IML any of a number of econometrics software packages (such as the TSP software I use) that can compute the panel data estimators, allowing for sample gaps, without requiring intricate matrix programming. Thus, it is possible and appropriate to dispense with the requirement that only a single block of data be used for each site.

The data quality implication of continuous data is sensible, though circumstantial. There is nothing about irregular reporting that necessarily implies that the reported data are of poor quality. The role of continuity is not, in my opinion, strong enough to justify a stricter continuity check.

Dr. Bradley's opinion about his freedom to adopt more stringent selection rules because of the large number of observations in his data sets is, arguably, the most controversial of the justifications for the strictness of the continuity checks. Dr. Bradley is correct that estimating his regressions on a subset of the available data (rather than the full data set) does not bias the parameter estimates, given that his sample selection rules are based on a priori rather than pretest criteria. However, while statistical theory indicates that it is permissible to estimate a regression model using less than the "full" set of observations described by the model—which is important, since any data set could contain some faulty observations—it does not suggest that it is desirable to do so. Dr. Bradley addresses this by stating that he considered the attendant loss of "efficiency" (i.e., increase in variance) of the estimates to be a reasonable trade-off for improved data quality. But since there is little or no presumption of material error in some of the lost observations, it is unclear whether the data quality improvement justified any efficiency loss.

Ultimately, Dr. Bradley's continuity check created more the appearance or risk of bias because of the large reduction in sample size than an actual bias, as Mr. Higgins pointed out (Docket No. R97–1, Tr. 33/18014). Still, it is a risk easily enough avoided with less stringent sample selection procedures. Therefore, I chose not to impose any continuity requirement at all, beyond that required by the specification of the labor demand models.

VI.D.3. "Productivity" scrub

Of Dr. Bradley's sample selection procedures, only the productivity check was clearly intended to identify and eliminate erroneous observations from the regression samples (Docket No. R97–1, USPS–T–14, at 32). Dr. Bradley based his productivity check on the observation that the extreme values of operation productivities (TPH per workhour) were too high or low to represent correctly reported data. In such cases, there would almost certainly be a flagrant error in either workhours or TPH.[52] The unusual feature of Dr. Bradley's procedure is that it worked by removing a small but fixed proportion of the observations (one

[52] Since MODS volumes and workhours are captured via separate systems, it is highly unlikely that flagrant errors with the same direction and magnitude would occur in both pieces of data.

percent from each tail of the productivity distribution) rather than applying criteria based on operational knowledge to identify and remove erroneous observations.

Removing a fixed proportion of the observations creates two potential problems. First, if fewer than two percent of the observations are clearly erroneous, Dr. Bradley's procedure will remove some observations that are merely unusual. Further, Dr. Bradley's approach to the continuity check magnifies the effect of the productivity check on the final sample by ensuring that at least some (and potentially all) observations before or after the gap left by the erroneous observation(s) are also removed. Second, to the extent that more than two percent of the observations are clearly erroneous, removing only the two percent of observations in the productivity tails leaves some number of erroneous observations in the regression sample. Interestingly, Dr. Bradley stated that he observed more "data problems" in the manual parcel and Priority Mail operations (Docket No. R97–1, Tr. 11/5284), but did not adjust his sample selection procedures for those operations accordingly.

A productivity check that removes a fixed fraction of the observations could be "excessive" for some operations by removing correct but unusual observations in higher quality data, but "ineffective" for operations with lower-quality data. The obvious solution to the problem is to apply some operational knowledge to the process and tailor the selection rules to the characteristics of the activities. Such an approach provides the greatest probability that the removed observations are actually erroneous.

VI.E. Summary of BY98 data and sample selection procedures

VI.E.1. The MODS data are of acceptable quality

One of the Commission's fundamental criticisms of Dr. Bradley's study was its dependence on MODS data (and PIRS data for BMC operations). I also rely on MODS for operation-specific workhours and piece handling volumes. The Commission stated that MODS had not been designed to produce data that met "econometric standards" and that the quality of the MODS data was, further, "far below the common standard" (PRC Op. R97–1, Vol. 1, at 82). It is difficult to evaluate the Commission's statements on MODS data quality relative to the econometrics literature, in part because there are no fixed "econometric standards" of which I am aware. However, after considering the interactions of the characteristics of the MODS data with the properties of the fixed-effects estimation procedures, I conclude that the MODS workhour and piece handling data for the sorting operations are of acceptable quality. The main relevant criticisms of the MODS data, which relate to the methods used to impute manual TPH, are flatly inapplicable to the mechanized and automated sorting operations. I find that the Inspection Service volume audit data used in Docket No. R97–1 to support the contention that there may be large, material errors in manual piece handlings are anecdotal and thus cannot support generalizations about the full MODS data set. On the other hand, statistical evidence developed in Docket No. R97–1 that is consistent with the absence of large, material errors in the manual data was incorrectly ignored, largely because of Dr. Neels's erroneous interpretation of it (see Section V.H, above).

To some extent, the lack of fixed standards reflects the fact that the econometrics "tool kit" contains many techniques to deal with common types of data problems. This is necessary in practice, because econometricians must accommodate an extremely broad spectrum of economic data quality. In my opinion, the Commission was overly pessimistic when it said, "Econometricians do not have very effective tools for identifying and correcting biases and inconsistencies caused by 'omitted variables' or 'errors-in-variables' unless the true error process is known" (PRC Op., R97–1, Vol. 1, at 82). I believe it would be more correct to say that the econometric tools for *efficient* estimation in the presence of omitted variables or errors-in-variables problems are not very effective unless the true data generating process is known. However, econometricians have many tools available for *consistent* estimation in the presence of various failures of "classical" assumptions. For instance, error components models such as fixed-effects and random-effects are completely effective at controlling for omitted factors associated with sites and/or time periods, when panel data are available. Recall that the site-specific dummy variables or "fixed effects," by construction, control for all of a site's "fixed" explanatory factors (see Sections II.C.2 and V.D).

The quality of real-world economic data ranges from the nearly pristine (e.g., financial markets data; some data on production processes collected by automated systems; well-designed and executed surveys) to the worse-than-useless (e.g., estimates of Soviet economic output). Most economic data fall in a broad middle range of intermediate quality. The fact that survey and experimental data are intended for subsequent statistical analysis does not automatically impart quality. It is well known that flaws in survey design can influence survey results. For example, economic measurement of capital often relies on imputations since services provided by assets such as buildings and equipment are harder to directly observe than labor inputs, and accounting methods sometimes fail to properly reflect the economic value of assets (e.g., a piece of equipment may be fully depreciated on a firm's books but still productive).[53]

Because of variations in MODS data collection methods and their interaction with the operating exigencies of the Postal Service, all of the MODS data should not be expected to be of equal quality. Piece handlings (TPF and TPH) in automated operations are collected automatically by the sorting equipment, and should be highly reliable barring possible transcription errors or technical difficulties with the automated end-of-run data transfers. Time clock data is also logged automatically, but it is not always efficient to have employees re-clock for every change of three-digit operation number. Therefore, workhours will tend to be more reliable when aggregated to the cost pool level. TPH in manual letter and flat sorting operations are subject to error from weight and

[53] Taking these factors into consideration, the Data Quality Study's suggestion that the Postal Service pursue an integrated analysis of mail processing labor and non-labor costs raises serious data quality issues separate from other mail processing modeling issues. See also Section VII.B.4.

downflow conversions, and thus will not tend to be of equal quality to their automation counterparts.

Based on a survey of the statistics literature, Hampel, et al., characterize data with gross errors of one to ten percent as "routine data" (Hampel, et al., p. 28), with "average quality" data commonly containing "a few percent gross errors" (Id., p. 64). My threshold and productivity checks are intended to identify the gross errors in the MODS data. Excluding the manual parcels and manual Priority Mail operations, these checks identify between 0.6 percent and 7.1 percent of the raw MODS observations as erroneous. The MODS data on those eight sorting operations appear, therefore, to be of approximately average quality—somewhat better for mechanized and automated operations than for manual operations. I summarize the effects of the sample selection rules in Table 3.

Table 3. Summary of Effect of Sample Selection Rules on Sample Size

Cost Pool	Non-missing	Threshold	Productivity	Minimum Obs	Lag Length (Regression N)
BCS	6885	6883	6780	6694	5391
			98.5%	97.2%	78.3%
OCR	6644	6639	6495	6394	5089
			97.8%	96.2%	76.6%
FSM	5442	5442	5424	5339	4357
			99.7%	98.1%	80.1%
LSM	5156	5150	5127	5014	3889
			99.4%	97.2%	75.4%
MANF	6914	6914	6033	5604	4427
			87.3%	81.1%	64.0%
MANL	6914	6914	6667	6511	5220
			96.4%	94.2%	75.5%
MANP	5835	5625	4545	3718	2853
			77.9%	63.7%	48.9%
Priority	5717	5644	4864	4017	3071
			85.1%	70.3%	53.7%
SPBS	2244	2239	2213	1966	1569
			98.6%	87.6%	69.9%
1CancMPP	6746	6718	6579	6483	5206
			97.5%	96.1%	77.2%

Percentages are of non-missing observations.

VI.E.2. MODS TPF edits

Since TPH is defined as TPF less rejects, in theory TPF should always exceed TPH. However, a number of observations have recorded TPH higher than TPF. A few of these observations appear to represent cases in which large "accounting adjustments" have been made to TPF (e.g., because the recorded value of TPF is negative). Unfortunately, the data do not allow adjustments to be separated from the other reported data. Additionally, some sites appear to have systematically under-reported TPF relative to TPH in FSM operations prior to mid-FY95.

I chose to "correct" observations with lower TPF than TPH by substituting TPH as the best available estimate of TPF. This assumes that, in the event of an anomaly, the TPH data are correct. I also tested two alternative procedures—using the TPF as recorded (which implicitly assumes the TPF data are more reliable than TPH), and eliminating observations showing the anomaly from the sample altogether. My results are not sensitive to the method used to treat the anomalous observations.

VI.E.3. Threshold check based on workhours

It is appropriate to exclude observations resulting from clocking errors or other sources of "noise" because they do not contribute useful information regarding the structure of production. Consequently, application of a threshold check to identify and exclude "noise" from the regression sample is justified. To avoid excluding data from sites that have small but regular operations, I sought to set very low thresholds. I also based the threshold on workhours, rather than TPH or TPF, to avoid the problem that exceeding a given threshold in terms of piece handlings requires many more workhours for some operations than for others. Thus, I set a threshold of forty workhours per quarter as a minimum below which "Function 1" sorting activities would not regularly operate.

A threshold of forty workhours per quarter threshold is very low relative to the typical size of the operations. For an observation of a particular activity not to pass the threshold, it could employ no more than the equivalent of one-twelfth of a full-time employee averaged over the course of the quarter. By comparison, the median observation passing the threshold for manual parcels (the smallest operation under study) reported 1,142 workhours per quarter—more than two and one-third full-time equivalent employees. As a result, I would expect the observations that do not pass the workhours threshold to be a byproduct of serious clocking error rather than the result of small but normal operations.

For the letter and flat sorting operations, which are present and in regular operation at most large mail processing facilities, virtually all observations pass the threshold check. However, for the manual parcel and manual Priority Mail operations, a non-negligible fraction of the observations—respectively, 3.8 percent and 1.4 percent—report fewer than forty hours per quarter. Examining the data, I found evidence that hours, volumes, or both are likely to be erroneous for most of the manual parcel and manual Priority Mail observations removed from the regression samples by the threshold check. The vast majority would not have passed the productivity checks since they imply impossibly high productivity levels. Therefore, I conclude that the observations excluded from the sample by the threshold check are actually erroneous.[54] See Table 4 for a comparison of the manual parcel and manual Priority Mail observations passing and not passing the threshold check.

Table 4. Median Workhours, TPH, and Productivity (TPH/Workhour) for Manual Parcels and Manual Priority Observations

	TPH (000)	Hours	Productivity
MANP > 40 Hr	389	1,142	294
MANP <= 40 Hr	201	10	22,833
Priority > 40 hr	528	2,324	212
Priority <= 40 Hr	54	13	4,177

VI.E.4. Productivity check based on operation-specific cutoffs

As I indicated in Section V.G, above, it is appropriate under good statistical practice to remove observations containing gross errors from the regression samples. Furthermore, since MODS volume and workhour data are recorded independently, and are thus unlikely to simultaneously contain errors of the same direction and magnitude, checking for anomalous productivity data is an effective means of identifying erroneous observations.

To separate erroneous observations from those that are merely anomalous, I set maximum and minimum productivity cutoffs specific to each sorting operation. For automated operations, I use information on machine

---

[54] I observed one small site that may have operated its manual parcels operation at a staffing level just below the threshold. However, it also reported implausibly high manual parcel productivities for several quarters, which would have rendered its data unusable.

throughputs and staffing standards provided by witness Kingsley to develop the maximum productivity cutoffs. In manual operations, I use typical productivity and my knowledge of the activities to judge the level at which productivity is unsustainably high. Erroneous data could also be manifest in unusually low levels of productivity. The most severe case of low productivity I discovered concerned one site that appeared to have reported its volume data in millions (rather than the usual thousands) of pieces for several quarters, so its productivities appeared to be approximately one thousand times lower than normal levels. I determine the minimum productivity cutoffs using productivity statistics and my judgment to identify levels of productivity below which the data are unlikely to be correct. The productivity cutoffs used to generate my recommended elasticities are listed in Table 5.

It is impossible to set unambiguous productivity cutoff points for the operations given the available data. In worker-paced operations, attainable productivity levels will depend critically on hard-to-quantify factors such as the quality of supervision and the complexity of sort schemes. Even in machine-paced operations, factors such as local variations in staffing practices mean there is not necessarily a maximum productivity. Similarly, there is no theoretical minimum productivity—only good management practice weighs against assigning employees to sorting operations when there is no mail to process.[55]

[55] The ready availability of output measures—piece handlings—for sorting operations would tend to make overstaffing of these operations highly visible and thus relatively easy for higher-level management to correct. Of course, some operations may need to be staffed at higher levels, e.g. for service reasons, than their volumes might otherwise appear to dictate.

Table 5. Minimum and Maximum Productivity Cutoffs (TPH/workhour) for Sorting Operations

Operation	Median (Interquartile range), full sample	Minimum	Maximum
OCR	7,041 (3,641)	500	15,000
LSM	1,373 (348)	150	2,800
BCS	7,821 (3,101)	500	22,500
FSM	742 (203)	150	1,500
MANF	554 (341)	100	1,000
MANL	771 (408)	100	1,400
SPBS	256 (125)	50	725
MANP	294 (382)	25	700
Priority	211 (231)	25	600
Cancellation	3,638 (1,937)	500	9,000

VI.E.5. Minimum observations requirement

My mail processing data set takes the form of an "unbalanced" panel. The term "unbalanced" refers to the property that the sites need not have the same number of observations.[56] Additionally, there need not be the same number of observations for every sorting operation present at a site. Not all data in an unbalanced panel are necessarily usable in the regression analysis. The data for some operations at some sites may be insufficient to reliably estimate all of the

[56] A panel data set with the same number of observations for each site is called "balanced."

parameters of the model. In extreme cases, it may be impossible to estimate some parameters of the model with the available data. As a result, it is necessary to impose a sample selection rule that ensures an adequate number of observations per site, or a minimum observations requirement.

I chose to require a minimum of eight observations, which need not be consecutive, per site. This requirement is somewhat stricter than necessary—the regressions could, in theory, be run with only two observations per site (assuming the presence of valid data for the required lags). However, I consider it good statistical practice to require more observations per site. Primarily, this is to improve the statistical quality of the estimates of the site-specific intercepts. By analogy, it is possible to compute an arithmetic mean from a sample of two data points, or to conduct a T-test with only one degree of freedom. However, it is rarely desirable to do so because of the relatively high sampling variance of the statistics. Confidence intervals for common statistical tests narrow rapidly as data are added and degrees of freedom increase.[57] The benefits of requiring more observations per site are not limited to the fixed-effects model. For example, the error-averaging benefit claimed by Dr. Neels for the between model presumes that there are sufficient observations to compute statistically reliable averages of each site's data.

[57] For instance, with one degree of freedom, to reject the null hypothesis that a regression coefficient is zero at the 5 percent significance level (two-tailed test) requires that the T-statistic exceed 12.706 in absolute value. With ten degrees of freedom, the critical value of the T-statistic for the same test drops to 2.228.

Since the choice of an eight observation requirement is discretionary, I tested the sensitivity of my results to the eight observation requirement by estimating the models with a less restrictive four observation requirement, and with a more restrictive nineteen observation minimum requirement that "sub-balances" the regression samples.[58] I provide detailed results in Appendix B.

I found that the less restrictive requirement of four observations had no material effect on the estimated elasticities. The elasticities resulting from the four observation requirement are all statistically indistinguishable from (within one standard deviation of) the corresponding results from the eight observation requirement. Furthermore, the composite variability of 75.8 percent in the four observation case is scarcely different from the 76.0 percent resulting from my preferred eight observation method. The more restrictive nineteen observation requirement has a larger relative effect on the elasticities in some operations. The largest effects occur in operations, such as LSM and Manual Priority, where the more restrictive selection requirement eliminates a large fraction of the potentially usable observations and facilities from the sample. Because of the large reductions in sample size, in my opinion, the sub-balanced samples are less likely to be representative of the entire spectrum of facilities with the operations than the more extensive samples provided by less restrictive requirements. Nevertheless, the composite variability of 75.0 percent is still quite

---

[58] Sub-balancing is selecting from an unbalanced panel a subset of observations that constitute a balanced panel. The nineteen observation requirement balances the data because nineteen observations is the maximum as well as the minimum number of observations per site.

close to the results from the less restrictive rules. Since the elasticities from both the less restrictive and more restrictive rules are actually somewhat lower, on balance, than those from the more extensive samples, I find no evidence that the eight observation minimum requirement imparts a downward bias on my estimated elasticities.

VII. Econometric Results

VII.A. Model specification and recommended results for MODS distribution operations

In this section, I present the estimating equations and major econometric results employed in the Postal Service's Base Year 1998 mail processing volume-variable cost analysis. I estimated labor demand functions for ten mail processing operation groups for which piece handling data are available from MODS. I employed two basic specifications of the labor demand equations. The variables included in the labor demand functions for all operations are piece handlings (current and the first four lags), the NWRS wage for the corresponding LDC, the facility capital index, possible deliveries as a measure of local network effects, a time trend, and seasonal (quarterly) dummy variables. For the six letter and flat sorting operations—BCS, FSM, LSM, OCR, manual letters, and manual flats—the labor demand models also include the manual ratio variable for the appropriate (letter or flat) shape as a measure of the degree of automation. The translog estimating equation also includes squared and interaction terms as indicated below. I estimated the models using quarterly data from quarter 2 of FY93 to quarter 4 of FY98, normalizing the quarter 4 observations to adjust for the extra accounting period. I computed the estimates using TSP software, version 4.4, published by TSP International.

The estimating equation for the six letter and flat sorting operations is:

$$
\begin{aligned}
\ln HRS_{it} = \beta_{1i} &+ (\alpha_1 + \gamma_1 L + \gamma_2 L^2 + \gamma_3 L^3 + \gamma_4 L^4)\ln TPH_{it} \\
&+ (\alpha_{11} + \gamma_{11} L + \gamma_{22} L^2 + \gamma_{33} L^3 + \gamma_{44} L^4)(\ln TPH_{it})^2 \\
&+ \alpha_2 \ln CAP_{it} + \alpha_{22}(\ln CAP_{it})^2 + \alpha_3 \ln DEL_{it} + \alpha_{22}(\ln DEL_{it})^2 \\
&+ \alpha_4 \ln WAGE_{it} + \alpha_{22}(\ln WAGE_{it})^2 + \alpha_5 TREND_{it} + \alpha_{55} TREND_{it}^{\;2} \\
&+ \alpha_6 \ln MANR_{it} + \alpha_{66}(\ln MANR_{it})^2 \\
&+ \alpha_{12} \ln TPH_{it} \ln CAP_{it} + \alpha_{13} \ln TPH_{it} \ln DEL_{it} + \alpha_{14} \ln TPH_{it} \ln WAGE_{it} \\
&+ \alpha_{15} \ln TPH_{it} \cdot TREND_{it} + \alpha_{16} \ln TPH_{it} \ln MANR_{it} \\
&+ \alpha_{23} \ln CAP_{it} \ln DEL_{it} + \alpha_{24} \ln CAP_{it} \ln WAGE_{it} + \alpha_{25} \ln CAP_{it} \cdot TREND_{it} \\
&+ \alpha_{26} \ln CAP_{it} \ln MANR_{it} \\
&+ \alpha_{34} \ln DEL_{it} \ln WAGE_{it} + \alpha_{35} \ln DEL_{it} \cdot TREND_{it} \\
&+ \alpha_{36} \ln DEL_{it} \ln MANR_{it} \\
&+ \alpha_{45} \ln WAGE_{it} \cdot TREND_{it} + \alpha_{46} \ln WAGE_{it} \ln MANR_{it} \\
&+ \alpha_{56} TREND_{it} \ln MANR_{it} \\
&+ \beta_2 QTR2_{it} + \beta_3 QTR3_{it} + \beta_4 QTR4_{it} \\
&+ \varepsilon_{it}.
\end{aligned}
$$

where the subscripts *i* and *t* refer to the site and time period, respectively, and *L* denotes the lag operator.[59]

[59] The lag operator is defined such that $L^s x_t = x_{t-s}$.

The estimating equation for the remaining operations—manual parcels, manual Priority, SPBS, and cancellation/meter prep—omits the terms associated with the manual ratio variable:

$$
\begin{aligned}
\ln HRS_{it} = {} & \beta_{1i} + (\alpha_1 + \gamma_1 L + \gamma_2 L^2 + \gamma_3 L^3 + \gamma_4 L^4) \ln TPH_{it} \\
& + (\alpha_{11} + \gamma_{11} L + \gamma_{22} L^2 + \gamma_{33} L^3 + \gamma_{44} L^4)(\ln TPH_{it})^2 \\
& + \alpha_2 \ln CAP_{it} + \alpha_{22} (\ln CAP_{it})^2 + \alpha_3 \ln DEL_{it} + \alpha_{22} (\ln DEL_{it})^2 \\
& + \alpha_4 \ln WAGE_{it} + \alpha_{22} (\ln WAGE_{it})^2 + \alpha_5 TREND_{it} + \alpha_{55} TREND_{it}^{\;2} \\
& + \alpha_{12} \ln TPH_{it} \ln CAP_{it} + \alpha_{13} \ln TPH_{it} \ln DEL_{it} + \alpha_{14} \ln TPH_{it} \ln WAGE_{it} \\
& + \alpha_{15} \ln TPH_{it} \cdot TREND_{it} \\
& + \alpha_{23} \ln CAP_{it} \ln DEL_{it} + \alpha_{24} \ln CAP_{it} \ln WAGE_{it} + \alpha_{25} \ln CAP_{it} \cdot TREND_{it} \\
& + \alpha_{34} \ln DEL_{it} \ln WAGE_{it} + \alpha_{35} \ln DEL_{it} \cdot TREND_{it} \\
& + \alpha_{45} \ln WAGE_{it} \cdot TREND_{it} \\
& + \beta_2 QTR2_{it} + \beta_3 QTR3_{it} + \beta_4 QTR4_{it} \\
& + \varepsilon_{it}.
\end{aligned}
$$

My preferred results are computed using a generalized least squares (GLS) procedure to allow the regression disturbances to exhibit First-order autocorrelation. The GLS procedure is a version of the "Baltagi-Li" autocorrelation adjustment (see Docket No. R97–1, USPS–T–14, at 50) modified to accommodate breaks in sites' regression samples.

As I discussed in Section V.F, I chose not to estimate the models on mean-centered data like Dr. Bradley, but rather explicitly evaluated the elasticity formulas to facilitate comparisons of results from different evaluation methods. In Tables 6 and 7, I present estimated elasticities rather than the estimated regression coefficients since the latter have no direct economic interpretation. I have provided output files containing the parameter estimates, as well as additional descriptive statistics, in the library reference LR–I–107.

Table 6. Principal results for letter and flat sorting operations, USPS Base Year method

Cost Pool	BCS	OCR	FSM	LSM	Manual Flats	Manual Letters
Output Elasticity or Volume-Variability Factor	.895 (.030)	.751 (.038)	.817 (.026)	.954 (.022)	.772 (.027)	.735 (.024)
Deliveries Elasticity	.250 (.046)	.333 (.062)	.223 (.037)	.039 (.045)	.313 (.043)	.462 (.040)
Wage Elasticity	-.826 (.052)	-.605 (.071)	-.613 (.041)	-.138 (.077)	-.232 (.060)	-.682 (.051)
Capital Elasticity	.024 (.019)	-.003 (.027)	.050 (.014)	.010 (.022)	.054 (.020)	.036 (.017)
Manual Ratio Elasticity	.071 (.015)	-.007 (.020)	-.047 (.011)	-.055 (.018)	-.032 (.028)	-.193 (.021)
Auto-correlation coefficient	.642	.701	.623	.558	.674	.693
Adjusted R-squared	.985	.970	.993	.991	.987	.990
N observations	5390	5088	4357	3889	4879	5499
N sites	297	289	235	273	277	299

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.

Table 7. Principal results for other operations with piece handling data, USPS Base Year method

Cost Pool	Manual Parcels	Manual Priority	SPBS	Cancellation & Meter Prep
Output Elasticity or Volume-Variability Factor	.522 (.028)	.522 (.025)	.641 (.045)	.549 (.037)
Deliveries Elasticity	.231 (.088)	.523 (.103)	.119 (.106)	.367 (.053)
Wage Elasticity	-.581 (.150)	-1.209 (.157)	-1.309 (.081)	-.580 (.086)
Capital Elasticity	.103 (.045)	.108 (.052)	.103 (.039)	.062 (.020)
Autocorrelation coefficient	.582	.500	.595	.669
Adjusted R-squared	.931	.940	.984	.982
N observations	3024	3241	1569	5235
N sites	181	200	94	290

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.

VII.B. Discussion of results

VII.B.1. General observations

My results show that costs in the mail processing operations for which I estimated volume-variability factors are, overall, much more responsive to changes in volume than total Postal Service costs or delivery labor, the other major field labor category. All of the estimated volume-variability factors in Tabl3 6 and Table 7 differ from 100 percent by at least two standard errors, and therefore are significantly less than 100 percent. There is also considerable variation from cost pool to cost pool, consistent with Mr. Degen's analysis as well as Dr. Bradley's Docket No. R97–1 results.

The composite volume-variable cost percentage for the mail processing operations with econometric volume-variability factors, 76.0 percent, is considerably higher than the overall percentage of volume-variable costs in the Postal Service's Base Year 1998 CRA, 60.4 percent. It is only slightly lower than the composite figure for Cost Segment 3, 77.4 percent, despite the application of the 100 percent volume-variability assumption to most other mail processing costs. It is almost fifty percent greater than the composite variability of delivery labor (Cost Segments 6, 7, 9,and 10), 51.3 percent, where 100 percent volume-variability is assumed for most in-office activities.[60]

---

[60] I obtained the volume-variable and total costs from the Base Year 1998 (Postal Service method) Cost Segments and Components Report, in witness Meehan's testimony (USPS–T–11).

In Appendix A, I present results using "all usable" observations in the regression sample, updating the exercise performed by Dr. Neels in Docket No. R97–1. The effect of including the data excluded by my sample selection rules in the regressions is to lower the overall variability for these operations by 8.7 percent, from 76.0 percent to 69.4 percent. Since many of the observations added to the samples between my preferred sample and the Appendix A samples are likely to be erroneous, I do not recommend the use of the Appendix A results.

VII.B.2. Specification tests unambiguously favor the fixed-effects model

Discussing the difference in volume-variability results among the pooled, between, and fixed-effects estimators, Dr. Neels stated, "...a good study should be like shooting elephants. It should be a really big target and easy to hit no matter how you do it" (Docket No. R97–1, Tr. 28/15786–7). I find Dr. Neels's implication that the results of a "good study" should be largely independent of methodology, runs completely counter to a huge body of theoretical and applied econometric research that underscores the importance of correct choice of methodology to produce accurate results. As such, there is no reason to be surprised that the results differ widely over estimators that embody such disparate statistical assumptions as do the pooled, cross-section, and fixed-effects estimators.

I would be more inclined towards Dr. Neels's view if it were necessary to choose among the estimators based on ambiguous results from specification tests. However, this is not remotely the case for the present study, nor was it the case for Dr. Bradley (see Docket No. R97–1, USPS–T–14, at 39–46). There is a clear sequence of specification tests to evaluate the applicability of the pooled model and, if the pooled model is rejected, of the fixed- versus random-effects formulations of the error-components model.[61] In this sequence, the pooled model—the most restrictive specification—is first tested against the error-components model. Then, if the pooled model is rejected, the choice is among the "fixed-effects" and "random-effects" specifications of the error-components model.

Testing the pooled model against the error-components model using a Lagrange multiplier test (see W. Greene, *Econometric Analysis,* 2/ed, MacMillan, New York, 1993, p. 450; Docket No. R97–1, USPS–T–14, at 41–43; Tr. 33/18021-18022), the pooled model's restrictions are decisively rejected for every cost pool. The significance levels are well below 1 percent for every cost pool except Manual Priority, where the pooled model is still strongly rejected at the 2.6 percent significance level. These results also weigh against cross-section estimators, including the between model, which also assume that the regression intercepts can be pooled over sites. Using a Hausman test of the random-effects formulation against fixed-effects (see C. Hsiao, *Analysis of Panel Data*, Cambridge University Press, 1986, p. 49), the more restrictive random-effects model is rejected in favor of fixed-effects at the 1 percent level or better in

[61] Recall from Section VI.E.1 that the "fixed-effects" and "random-effects" models are types of error-components models.

every case but SPBS, where the significance level is 2.33 percent. See Table 8. Therefore, the statistical case in favor of the fixed-effects model is clear. To favor a less general pooled or cross-section estimator over fixed-effects is to ignore unambiguous specification test results to the contrary.

Table 8. Significance levels (P-values) for specification tests

Cost Pool	Pooled vs. Error components	Fixed effects vs. Random effects
BCS	0.001	0.0004
OCR	< 0.0005*	< 0.00005*
Manual Flats	< 0.0005*	< 0.00005*
Manual Letters	< 0.0005*	< 0.00005*
FSM	< 0.0005*	< 0.00005*
LSM	< 0.0005*	0.0001
SPBS	< 0.0005*	< 0.0233
Manual Parcels	< 0.0005*	< 0.00005*
Manual Priority	0.026	< 0.00005*
Cancellation/Meter Prep	< 0.0005*	< 0.00005*

* All reported digits of P-value are zero.

VII.B.3. The network has a significant impact on mail processing costs

The local network, measured by possible deliveries, has a significant effect on costs in most of the mail processing operations.[62] For the most part,

[62] The exception is LSM, where the deliveries elasticity is positive but small and not statistically significant.

the deliveries elasticities are considerably lower than the volume-variability factors. The exception is the manual Priority Mail cost pool, where the volume-variability factor and the deliveries elasticity have approximately equal magnitude. Since the Priority Mail operation has relatively large setup and takedown activities according to Mr. Degen's description (USPS–T–16, at 44), and staffing needs would tend to be relatively high to meet Priority Mail's service standards, this result should not be surprising.

The deliveries elasticities are the quantitative reflection of Mr. Degen's fundamental observation that characteristics of the service area play a significant role in determining mail processing costs (Id., at 6). The econometric results demonstrate that the costs of a "large" mail processing operation are due not only to the volume of mail processed therein, but also due to the extent of the network being served.[63] The significance of the distinction between the volume and the network effect for postal costing is that *the deliveries elasticities, the contributions of the network to the costs of mail processing operations, are not causally attributable to the subclasses of mail.*

---

[63] This closely mirrors findings for other network industries. See, e.g., D. Caves, L. Christensen, M. Tretheway 1984; D. Caves, L. Christensen, M. Tretheway, and R. Windle, "Network Effects and the Measurement of Returns to Scale and Density for U.S. Railroads," in A. F. Daugherty, ed., *Analytical Studies in Transport Economics*, Cambridge University Press, 1985.

VII.B.4. Comparison to Dr. Bradley's results

A striking feature of my results is their general consonance with the corresponding results of Dr. Bradley's study from Docket No. R97–1. I compare my results to Dr. Bradley's in Table 9 below. Certainly, there are large changes

Table 9. Comparison of Postal Service BY1996 and BY1998 volume-variability factors

Cost Pool	BY 1996 Variability (Docket No. R97–1, USPS–T–14)	BY 1998 Variability	Percent difference – BY98 vs. BY96
BCS	.945	.895	-5.3%
OCR	.786	.751	-4.5%
Manual Flats	.866	.772	-10.9%
Manual Letters	.797	.735	-7.8%
FSM	.918	.817	-11%
LSM	.905	.954	5.4%
SPBS	.552[64]	.641	16.1%
Manual Parcels	.395	.522	32.2%
Manual Priority	.448	.522	16.5%
Cancellation and Meter Prep	.654	.549	-16.1%
Composite	.810	.760	-6.2%

for individual cost pools. Insofar as there are a number of material differences in the two analyses, notably my inclusion of additional economic variables in the

[64] Volume-variable cost percentage for combined SPBS – Priority and SPBS – Non-Priority cost pools. See Docket No. R97–1, USPS–T–12, at 15 [Table 4].

models and the differences in the time periods covered by the panel data sets, this is to be expected. Nonetheless, the basic ordinal relationships among the estimated variabilities are quite similar. I show that operations with relatively high variabilities in Dr. Bradley's study (e.g., BCS; letter and flat sorting as a group) still exhibit relatively high degrees of volume-variability. While there are large upward revisions to the manual parcel and Priority variabilities, due largely to the application of tighter sample selection rules, those cost pools still have relatively low variabilities. Overall, the composite degree of variability for the cost pools I studied differs from the Docket No. R97–1 figure by only 6.2 percent.

Since the additional explanatory variables—particularly wages and network variables—are statistically significant, my results indicate that Dr. Bradley's Docket No. R97–1 mail processing models for the operations I studied were underspecified. As a result, Dr. Bradley's results appear to exhibit omitted-variables bias to some degree. However, since the revised variabilities accounting for these factors are lower, contrary to the expectations set forth in the Commission's Docket No. R97–1 analysis, the direction of the omitted variables biases in Dr. Bradley's results were mainly upwards, not downwards. Thus, Dr. Bradley's estimates were much closer to the "true" volume-variability factors than the 100 percent volume-variability assumption or the pooled and cross-section results favored by, respectively, Dr. Smith and Dr. Neels that turn out to have greater upward biases.

VII.B.5. Relationship to Data Quality Study discussion of mail processing

The Data Quality Study's discussion of mail processing presents the suggestion that the 100 percent volume-variability assumption may have been warranted in the past, but that lower volume-variability factors may be applicable to the present mail processing system (see, e.g., Summary Report, at 76; Technical Report #4, at 41–42). As I discussed in Section I.D, a full reading of the Docket No. R71–1 record shows that there was no reliable statistical evidence to justify the 100 percent volume-variability assumption, apart from a simple regression analysis understood to be inadequate and not relied upon by its authors. Therefore, I consider the actual volume-variability of mail processing labor costs as of the R71–1 era's cost studies to be an open question.

The Data Quality Study views the shift from manual to automated sorting technology as a factor likely to cause a drop in the overall degree of volume-variability for mail processing. However, both Dr. Bradley and I have found higher measured volume-variability factors for the automated sorting operations than for manual sorting operations. The problem seems to be the Data Quality Study's authors' belief that automated operations "demand less piece handling" than manual operations, or are "batch driven" (Technical Report #4, at 43).[65] As

---

[65] The Data Quality Study's further implication appears to be that "batch driven" activities have low or zero volume-variability. This is a conjecture that can, in principle, be put to the test. A variability conjecture for batch driven activities can err on the side of low volume-variability by incorrectly assuming that the number of batches (per unit of time) is immutable. It can also err in favor of high volume-variability by improperly assuming proportionality between batches and volumes, thus denying the obvious possibility that marginal mail volumes could sometimes hitch a "free ride" with existing batches. In general, all the researcher can say is

Mr. Degen's and Ms. Kingsley's descriptions correctly indicate, pieces of mail are sorted no less individually on automated equipment than at a manual case. Furthermore, a worker at a manual case can vary the degree of effort based on the volume of mail to be processed while a piece of automation equipment cannot do so. As a result, there is no reason to suppose that direct physical contact with the mail necessarily results in manual operations having higher volume-variability factors than automation operations.[66]

The Data Quality Study also advocates integrating the cost analyses for mail processing labor, capital, and other inputs (Summary Report, p. 39–40). I believe this may be a useful indication of the long term direction of the economic analysis of Postal Service costs. However, much additional study and, potentially, data collection would be required to implement the proposed change. Since capital costing is one of the most contentious areas of applied economics, I consider it unlikely that the proposed change could be implemented without great controversy. Additionally, the CRA would need considerable reorganization, since the costs of most major activities (but particularly delivery and mail processing) of the Postal Service are divided among several current cost segments. Since it is economically valid either to estimate the full factor demand system, or just a portion (such as labor demand), I find the implementation of an

---

that the costs for such activities are volume-variable to the extent that additional volumes cause additional batches.

[66] Witness Moden made this point in Docket No. R97–1. See USPS–T–4, at 19.

integrated cost analysis unnecessary as a prerequisite for any methodological improvements.

VII.C. Results from alternative estimation methods

In Appendixes E, F, and G, I present the principal econometric results from three alternative estimation methods—respectively, the cross-section "between" model, the pooled model, and the random-effects model. I do not recommend the use of the results of any of these models because, as I demonstrate in Section VII.B.2, they all incorporate statistical restrictions that are decisively rejected in specification tests and therefore are biased. The relative performance of estimates from different methods can be examined by comparing the distributions of the estimated elasticities.

I compared the variances and distribution statistics for the estimated elasticities for the fixed-effects model and the "between" model. I found evidence that the between model's estimates are far more seriously affected by near-multicollinearity in the data. This is reflected both in higher standard errors of the aggregate elasticities and the smaller numbers of estimated coefficients that are "significant" in individual T-tests, relative to the fixed-effects model. I also found that the between model produces an extremely broad range of elasticity estimates for the individual data points, including negative and extremely large positive elasticities for extreme values. I believe the possibility that Postal Service sorting operations actually exhibit the degree of diversity in cost-output relationships, including the economically perverse result that an increase in handlings could sometimes reduce workhours, implied by the between model is infinitesimal at most. Therefore, I find that the results of the between model, even augmented with additional control variables, are completely unreliable. The fixed-effects model, in contrast, produces elasticities in a much narrower range, and with far more reasonable values at the extremes of the distributions. Table 10 provides a comparison of results for the manual letter and manual flat operations (similar results obtain for other cost pools; see LR–I–107 for additional results).

Table 10. Comparison of Selected Diagnostic Statistics for the Fixed-Effects and Between Models, Manual Letters, and Manual Flats cost pools

	Manual Letters		Manual Flats	
	Fixed-Effects	Between	Fixed-Effects	Between
Output Elasticity (arithmetic mean method)	.735	.901	.772	.543
Standard error	.024	.227	.027	.190
Number of "significant" slope coefficients (90% confidence level)	27	13	21	14
Median elasticity	.699	.871	.740	.596
Interquartile range of elasticities	.099	.247	.062	.638
Minimum elasticity	.266	-0.448	.494	-2.039
Maximum elasticity	1.197	1.889	1.067	2.530

VIII. Volume-Variability Assumptions for Other Cost Pools

VIII.A. The Postal Service's Base Year method adopts IOCS-based volume-variable costs for "non-measured" cost pools, but with significant reservations

The Postal Service's Base Year CRA employs volume-variability factors based on the traditional IOCS-based methods, rather than updated or extended versions of Dr. Bradley's models or assumptions, for several groups of cost pools. These are MODS allied labor, Registry, and BMC operations, for which Dr. Bradley provided econometric elasticities in Docket No. R97–1; and miscellaneous MODS and all non-MODS (and non-BMC) operations, for which Dr. Bradley provided alternative variability assumptions, generally based on analogies with cost pools with estimated elasticities.

The Postal Service's Base Year method determines the volume-variability factors for these cost pools as the ratio of the dollar-weighted IOCS tallies with "variable mail processing" activity codes to the total dollar-weighted tallies associated with the cost pool, excluding the "variable overhead" codes 6521-6523.[67] Details of the calculations are provided in the testimony of witness Van-Ty-Smith (USPS–T–17). The Postal Service's method differs from the Commission's accepted method in that the cost pools and tallies have not been

[67] The effect of excluding variable overhead is to make the time associated with breaks, empty equipment, and clocking in/out volume-variable to the same extent as the non-overhead activities in the same cost pool.

adjusted for so-called "migrated" costs at MODS offices.[68] Therefore, the Postal Service's implementation of the IOCS-based volume-variability method classifies a tally as volume-variable or "fixed" independent of the subpart of IOCS question 18 used to identify the employee's activity.

My explanation of the Postal Service's decision to use volume-variability factors based on the traditional IOCS activity code classification should not be construed as an endorsement of the traditional method on its economic merits. The Postal Service's Base Year CRA employs traditional variability methods for the cost pools mentioned above primarily because those methods are the status quo in the Commission's accepted methodology. However, an additional consideration in the Postal Service's decision was the degree to which research conducted in response to the Commission's Docket No. R97–1 Opinion was able to address the Commission's claimed "disqualifying defects" that led the Commission to reject Dr. Bradley's methodology and elasticity estimates.

VIII.B. Status of research on volume-variability factors for other operations

VIII.B.1. Non-MODS operations and MODS operations "without piece handlings" (except allied labor)

Dr. Bradley specified alternative assumptions (to the IOCS-based status quo) rather than econometric estimates for the non-MODS offices and a number

---

[68] This is consistent with witness Degen's recommendation that clerk and mail handler costs at MODS offices be partitioned into the mail processing, window service, and administrative components of Cost Segment 3 using the MODS operation number rather than the IOCS sampled activity.

of MODS cost pools because of the lack of data on workload drivers for those operations. No new sources of workload data have become available in the interim to permit estimation of elasticities for these cost pools. While witness Degen's testimony does not directly address these operations, many of the factors he identifies as consistent with lower volume-variability factors for Function 1 operations are also present in the analogous Function 4 and non-MODS operations. For many of the support-type operations the IOCS-based variability method already produces relatively low implicit variabilities.[69] However, I believe Mr. Degen's description of the structure of mail processing costs is also suggestive of a potential disconnection between the IOCS method of parsing tallies into fixed and variable categories and the real cost drivers for support operations, which are workhours and/or workload in the supported operations. The Postal Service has no quantitative evidence to support or refute either Dr. Bradley's assumptions from Docket No. R97–1 or the IOCS-based method. Absent evidence to overturn the status quo, the Postal Service decided to use the IOCS-based analysis for its Base Year CRA in this case.

I expect that the Postal Service will be able to provide quantitative evidence to bolster the quantitative analysis for some of these operations in the future. In the case of MODS mail processing support operations, data on the workhours and/or workload in the supported operations already exist. Therefore,

[69] A relatively high percentage of the tallies with activity codes traditionally classified as "fixed" mail processing fall in these cost pools. As a result, some cost pools actually have lower (implicit) volume-variability factors in the IOCS-based method than from Dr. Bradley's assumptions.

it is possible in principle to put them to use to estimate elasticities for the support operations and refine the piggyback assumptions. In the case of distribution operations at MODS stations and branches and at non-MODS post offices, it may also be possible to collect piece-handling data in the future, either through existing infrastructure (MODS) or through future data collection initiatives.

VIII.B.2. BMC operations

Time and resource constraints prevented the Postal Service from updating Dr. Bradley's BMC models, a process which would include examining what I believe to be a central issue underlying the Commission's rejection of the BMC models, the properties and quality of the PIRS data. It would also be desirable to investigate the availability of data to more fully specify the labor demand models for BMC operations, in a manner similar to the treatment of the MODS piece sorting operations. Insofar as the Postal Service does not have additional evidence that might persuade the Commission to adopt Dr. Bradley's models and results, it was decided to use the previously accepted variability method for the BMCs.

Nonetheless, I believe Dr. Bradley's efforts, though flawed in some respects, provide the best available estimates of elasticities for BMC operations. Extrapolating from the effects of the methodological changes on the MODS elasticities, I believe Dr. Bradley's models represent a much more accurate method for estimating the volume-variable costs in BMC operations than the IOCS-based method. I do not consider the PIRS data problems as necessarily disqualifying. The generic statistical issues arising from the use of "noisy" workload data have been addressed above in the context of the MODS piece handling volumes. However, the conversion of BMC workload into standard units—i.e., Total Equivalent Pieces (TEP)—provides an additional potential source of error in workload measurement that merits separate study. I cannot rule out the possibility that the PIRS data issues are serious, but I note that the PIRS workload data would have to be so noisy as to be useless in order for the IOCS-based method not to significantly overstate the BMC volume-variable costs relative to Dr. Bradley's methods.

VIII.B.3. MODS allied labor operations

Witness Degen's analysis of the characteristics of the allied operations suggests that the operational basis for reduced volume-variability factors (relative to the IOCS-based method) is at least as strong for allied operations as for sorting operations.[70] In addition, I was able to investigate Dr. Bradley's methodology for the MODS cost pools further, and provide some corroboration of his results. However, data limitations leave some important quantitative volume-variability questions unanswered. As a result, the Postal Service decided to continue using the IOCS-based analysis to determine volume-variable costs for the current filing. However, I would expect MODS allied labor to be a top priority for further study, due to the large accrued costs in the operations and their

[70] Mr. Degen's analysis also indicates that allied operations should be expected to have lower volume-variability factors than sorting operations.

importance in the development of marginal and incremental costs for Periodicals and presorted Standard A mail.

In Docket No. R97–1, Dr. Bradley specified models for MODS allied labor that used piece handling volumes at sorting operations as cost drivers, in lieu of data on handlings of "items" (bundles, trays, sacks and pallets) and rolling containers that are the focus of allied labor activities. The use of volumes from sorting operations as allied labor cost drivers has an operational foundation, since one purpose of the allied labor operations is to prepare mail for sorting in the facility, and to prepare mail that has been sorted for shipment to other facilities. However, as several witnesses indicated in Docket No. R97–1, a portion of the volume-related workload in allied operations consists of handling mail that bypasses piece sorting operations in the facility and therefore does not generate TPF, TPH, or FHP (see, e.g., Docket No. R97–1, USPS–T–14, at 2; Tr. 34/18226). Additionally, descriptions of platform activities have long recognized that vehicle arrivals and departures are also drivers of platform workload. Thus, a fully specified factor requirements models for allied labor operations could include variables representing several cost drivers in addition to the piece handling volumes from sorting operations used by Dr. Bradley. Indeed, Dr. Bradley recognized this in his Docket No. R97–1 testimony.[71] Such data, if it can be reliably collected and incorporated into cost models for the allied

[71] "Although it would be preferable to have a cost driver that directly measures workload in the allied activity, a good first attempt at measuring the variability of allied hours can be made by testing the assumption that allied hours are caused by the piece handlings in direct activities." Docket No. R97–1, USPS–T–14, at 2 (footnote omitted).

operations, could greatly facilitate the resolution of longstanding controversies that focus on the causal relationship between workshared mail classes (particularly Periodicals and presorted Standard A) and the costs incurred in allied labor operations.

As part of my study I began the effort of examining other data that might provide improved workload measures for MODS allied operations. At a basic level, my efforts were limited by the lack of ongoing Postal Service data systems to collect information on item and container handlings among other potential drivers of workload in the allied operations. However, I had access to data on crossdocked containers collected in a Christensen Associates survey of platform operations. I also estimated models using ODIS destinating volumes as cost drivers, as a means of capturing the bypass workload. Finally, I explored the applicability of data on the number of truck arrivals and departures from the TIMES system for use as a platform cost driver. In general, the results from models enhanced with these additional data indicated that Dr. Bradley's "proxy" cost drivers—the volumes from piece sorting operations—still provided the bulk of the explanatory power. In my opinion, these results likely reflect limitations of the available data for representing the workload components of main interest. Further, while TIMES is an ambitiously conceived system that may someday provide a great deal of data for use in the analysis of platform costs (among other things), it is also a relatively new system whose quality, for purposes such as my study, is unknown.[72]

[72] The recent Data Quality Study's report advocated investigating the applicability of TIMES data for the allied labor analysis. TIMES was not among the data systems reviewed for the Data Quality Study, and the Study's report therefore did not comment on its reliability.

Appendix A. Results from including "all usable" observations in the regression samples

Table A–1. Cost pool and composite volume-variability factors from all usable observations in sample

Cost Pool	Preferred Sample (USPS BY98)	"All usable" observations	Difference in number of obs.
BCS	89.5%	78.7%	244
OCR	75.1%	73.8%	332
FSM	81.7%	66.1%	70
LSM	95.4%	104.3%	81
Manual Flats	77.2%	72.7%	787
Manual Letters	73.5%	66.8%	162
Manual Parcels	52.2%	42.5%	1479
Manual Priority	52.2%	61.2%	1144
SPBS	64.1%	51.7%	110
Cancellation & Meter Prep	54.9%	41.6%	256
Composite Variability	76.0%	69.4%	

Appendix B. Results based on alternative minimum observation requirements

Table B–1. Volume-variability factors from four, eight, and nineteen observation minimums

Cost Pool	Eight observation minimum	Four observation minimum	Nineteen observation minimum
BCS	.895	.903	.924
OCR	.751	.743	.725
FSM	.817	.814	.807
LSM	.954	.953	.802
Manual Flats	.772	.771	.743
Manual Letters	.735	.735	.725
Manual Parcels	.522	.518	.531
Manual Priority	.522	.529	.400
SPBS	.641	.600	.636
Cancellation & Meter Prep	.549	.551	.576
Composite Variability	.760	.758	.750

Table B–2. Observations (top number) and sites (bottom number) from four, eight, and nineteen observation minimums

Cost Pool	Eight observation minimum	Four observation minimum	Nineteen observation minimum
BCS	5390 297	5408 300	4635 244
OCR	5088 289	5121 295	4047 213
FSM	4357 235	4373 238	4085 215
LSM	3889 273	3931 280	684 36
Manual Flats	4879 277	4921 284	3990 210
Manual Letters	5499 299	5506 300	4977 262
Manual Parcels	3024 181	3197 212	2090 110
Manual Priority	3241 200	3398 230	1881 99
SPBS	1569 94	1599 99	893 47
Cancellation & Meter Prep	5235 290	5265 295	4389 231

Appendix C. Derivation of the elasticities of the manual ratio with respect to piece handlings and volumes

Let D_m denote manual piece handlings and D_a denote automated and mechanized piece handlings. Then the manual ratio is

$$MANR = \frac{D_m}{D_m + D_a}.$$

The Commission computed the following derivatives of the manual ratio with respect to piece handlings (PRC Op., R97–1, Vol. 2, Appendix F, at 39):

$$\frac{\partial MANR}{\partial D_m} = \frac{D_a}{(D_m + D_a)^2} = \frac{1 - MANR}{D_m + D_a}$$

$$\frac{\partial MANR}{\partial D_a} = -\frac{D_m}{(D_m + D_a)^2} = -\frac{MANR}{D_m + D_a}.$$

The above derivatives are nonzero (as long as both manual and automated handlings are both positive) and have opposite signs.

The derivatives of the manual ratio with respect to piece handlings are not sufficient to quantify the potential contribution of the manual ratio to the overall degree of volume-variability. Rather, the "manual ratio effect" involves three types of terms: the elasticity of workhours with respect to the manual ratio, the elasticities of the manual ratio with respect to piece handlings, and the elasticities of piece handlings with respect to volumes. Mathematically, the manual ratio effect with respect to the volume of subclass j is:

$$\mu_j = \frac{\partial \ln HRS}{\partial \ln MANR} \cdot \left(\frac{\partial \ln MANR}{\partial \ln D_m} \cdot \frac{\partial \ln D_m}{\partial \ln V_j} + \frac{\partial \ln MANR}{\partial \ln D_a} \cdot \frac{\partial \ln D_a}{\partial \ln V_j} \right).$$

Extending the Commission's results, the elasticities of the manual ratio with respect to piece handlings are:

$$\frac{\partial \ln MANR}{\partial \ln D_m} = \frac{D_m}{MANR} \cdot \frac{\partial MANR}{\partial D_m} = \frac{D_m(1-MANR)}{MANR(D_m + D_a)} = 1 - MANR$$

$$\frac{\partial \ln MANR}{\partial \ln D_a} = \frac{D_a}{MANR} \cdot \frac{\partial MANR}{\partial D_a} = -\frac{D_a \cdot MANR}{MANR(D_m + D_a)} = -(1 - MANR).$$

The elasticities of the manual ratio with respect to manual and automated piece handlings have equal magnitudes but opposite signs.

Under the "proportionality assumption," the elasticities of piece handlings with respect to the volume of subclass *j* is approximately equal to subclass *j*'s share of piece handlings, or the distribution key share. That is, given

$$D_{(a,m)} = \sum_{i=1}^{N} D_{(a,m),j} = \sum_{i=1}^{N} \alpha_i^{(a,m)} V_i ,$$

where the summation is over subclasses, the elasticity of piece handlings with respect to volume is:

$$\frac{\partial \ln D_{(a,m)}}{\partial \ln V_j} = \frac{V_j}{D_{(a,m)}} \cdot \frac{\partial D_{(a,m)}}{\partial V_j} = \frac{\alpha_j^{(a,m)} V_j}{D_{(a,m)}} = \frac{D_{(a,m),j}}{D_{(a,m)}} \equiv \delta_j^{(a,m)} .$$

As the Commission observed (Id., at 41), subclasses will differ in their "contributions" to manual and automated piece handlings because of differences in automation compatibility, preparation, and other characteristics. In fact, the IOCS-based distribution keys for the letter and flat sorting operations confirm that the distribution key shares in manual and automated operations differ for most subclasses.

Combining results, we obtain:

$$\begin{aligned}\mu_j &= \varepsilon_{HRS,MANR}\cdot\left[(MANR-1)\cdot\delta_j^a+(1-MANR)\cdot\delta_j^m\right]\\ &= \varepsilon_{HRS,MANR}\cdot(1-MANR)\cdot(\delta_j^m-\delta_j^a)\end{aligned}.$$

Therefore, there is a manual ratio effect for subclass *j* unless the subclass accounts for an equal share of handlings in both the manual and automated operations. However, the manual ratio effect is a "zero-sum game"—it may affect the costs for an individual subclass, but the net effect summed over all subclasses is zero:

$$\begin{aligned}\sum_{j=1}^{N}\mu_j &= \varepsilon_{HRS,MANR}\cdot(1-MANR)\cdot\left(\sum_{j=1}^{N}\delta_j^m-\sum_{j=1}^{N}\delta_j^a\right)\\ &= \varepsilon_{HRS,MANR}\cdot(1-MANR)\cdot(1-1)=0\end{aligned}.$$

Thus, the overall degree of volume-variability for the cost pool is not underestimated by neglecting the manual ratio effect.

The manual ratio effect is measurable using the elasticities of workhours with respect to the manual ratio for the letter and flat sorting operations, average values of the ratios themselves, and distribution key data. Since the elasticities of workhours with respect to the manual ratio are relatively low for the letter and flat sorting cost pools, it is not obvious that the manual ratio effect would be particularly large. I estimated the manual ratio effects for each CRA subclass (using the spreadsheet TableC1.xls provided in LR–I–107), expressed each as a percentage of the total volume-variable cost for MODS offices, and compared the size of the effects to the coefficients of variation provided by witness Ramage (USPS–T–2). See Table C–1 for the results. The estimated manual ratio effect is smaller than the coefficient of variation, and thus is within the sampling error of the volume-variable cost estimates, for all subclasses except First-Class Presort (2.5 percent) and First-Class Cards (-4.7 percent). The effect for First-Class Cards is outside an approximate 80 percent confidence interval but inside the 90 percent confidence interval for the cost estimate and is "borderline" significant. While the preferable treatment of the manual ratio is, as I explain in Section II.C.2, as a non-volume technology control variable, the effect on measured subclass volume-variable costs would be relatively small if it were to be treated as "volume-variable."

Table C–1. Comparison of Manual Ratio Effect and Sampling Error of MODS Volume-Variable Cost Estimates by Subclass

Subclass	Manual Ratio Effect (percentage of MODS volume-variable cost)	Coefficient of Variation (CV) – USPS–T–2	Percentage Manual Ratio Effect, Fraction of CV
1st L&P	-0.3%	0.5%	-0.71
1PreL	2.5%	1.2%	2.03
1Cds	-4.7%	3.1%	-1.48
1PreC	-2.7%	7.1%	-0.38
Priority	-0.2%	1.3%	-0.16
Express	-0.2%	3.4%	-0.06
Mailgrams	-14.4%	65.5%	-0.22
2IC	-1.9%	12.1%	-0.16
2Reg	-1.3%	1.7%	-0.74
2NP	-1.0%	4.4%	-0.24
2CL	-0.4%	18.8%	-0.02
3SP	0.0%	5.0%	0.00
3BRCRT	1.1%	2.6%	0.42
3BRO	0.3%	0.9%	0.36
3NPCRT	1.1%	7.2%	0.15
3NPO	-0.6%	2.3%	-0.25
4ZPP	-0.1%	3.1%	-0.04
4BPM	-0.3%	4.9%	-0.06
4SPC	0.0%	6.1%	0.00
4LIB	-0.4%	13.2%	-0.03
USPS	-1.4%	3.8%	-0.36
Free	-1.1%	13.2%	-0.08
Intl	-0.2%	2.3%	-0.07

Appendix D. Algebraic and econometric results pertaining to alternative elasticity aggregation methods

In this appendix, I demonstrate the algebraic equivalence of "average of the variabilities" methods with the geometric mean methods, given the translog form of the labor demand models.

<u>Preliminaries</u>

The (unweighted) geometric mean is defined as follows:

$$\langle X \rangle = \left(X_1 \cdot X_2 \cdot \ldots \cdot X_N\right)^{1/N} = \left(\prod_{i=1}^{N} X_i\right)^{1/N},\ X_i > 0.$$

I define the weighted geometric mean as:

$$\langle X \rangle_W = X_1^{\sigma_1} \cdot X_2^{\sigma_2} \cdot \ldots \cdot X_N^{\sigma_N} = \prod_{i=1}^{N} X_i^{\sigma_i},\ X_i, \sigma_i > 0,\ \sum_{i=1}^{N} \sigma_i = 1,$$

where the terms σ_i are weighting factors. If the weights are uniformly equal to 1/N, the weighted and unweighted geometric means are identical. A result that I use below is that the natural logarithm of the geometric mean of X is equal to the arithmetic mean of the natural logarithm of X:

$$\ln\langle X \rangle = \ln\left(\left(\prod_{i=1}^{N} X_i\right)^{1/N}\right) = \frac{1}{N}\ln\left(\prod_{i=1}^{N} X_i\right) = \frac{1}{N}\sum_{i=1}^{N} \ln X_i = \overline{\ln X_i}.$$

A comparable result holds for the weighted geometric mean, i.e.:

$$\ln\langle X \rangle_W = \sum_{i=1}^{N} \sigma_i \ln X_i.$$

Recall from Section V.F that the elasticity functions derived from the translog labor demand equation (with K explanatory variables $x_1, \ldots x_K$) have the form:

$$\varepsilon_j = \partial \ln hrs / \partial \ln x_j = \alpha_j + \sum_{k=1}^{K} \alpha_{jk} \ln x_k.$$

Finally, consistent (or unbiased) estimates of ej can be obtained by substituting appropriate estimates for the α parameters:

$$\hat{\varepsilon}_j = \hat{\alpha}_j + \sum_{k=1}^{K} \hat{\alpha}_{jk} \ln x_k \, .$$

Note that the value of the elasticity depends on the values chosen for the explanatory variables $x_1, \ldots x_K$.

<u>"Average-of-the-variabilities" methods and the geometric mean</u>

The simplest average variability method is to oompute the arithmetic mean of the estimated elasticities of the N observations in the sample, that is:

$$\bar{\varepsilon} = \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_i \, .$$

Witness Andrew proposed this method in Docket No. R90–1. The volume f is predicted using the estimated parameters of the labor demand function with:

$$\hat{\varepsilon}_j = \hat{\alpha}_j + \sum_{k=1}^{K} \hat{\alpha}_{jk} \ln x_k \, .$$

Combining the above two formulas results in:

$$\begin{aligned}
\overline{\hat{\varepsilon}_j} &= \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\alpha}_j + \sum_{k=1}^{K} \hat{\alpha}_{jk} \ln x_{ki} \right) \\
&= \hat{\alpha}_j + \sum_{k=1}^{K} \hat{\alpha}_{jk} \frac{1}{N} \sum_{i=1}^{N} \ln x_{ki} \\
&= \hat{\alpha}_j + \sum_{k=1}^{K} \hat{\alpha}_{jk} \overline{\ln x_k} \\
&= \hat{\alpha}_j + \sum_{k=1}^{K} \hat{\alpha}_{jk} \ln \langle x_k \rangle .
\end{aligned}$$

That is, averaging the estimated elasticities is equivalent to evaluating the elasticity function at the geometric mean of the explanatory variables. A similar result obtains for a weighted average elasticity—that is, taking a weighted average of the elasticities is equivalent to evaluating the elasticity function at the weighted geometric mean of the explanatory variables (using the same weights):

$$\begin{aligned}\left(\hat{\bar{\varepsilon}}_j\right)_W &= \sum\nolimits_{i=1}^{N} \sigma_i \hat{\varepsilon}_i \\ &= \hat{\alpha}_j + \sum\nolimits_{k=1}^{K} \hat{\alpha}_{jk} \sum\nolimits_{i=1}^{N} \sigma_i \ln x_{ki} \\ &= \hat{\alpha}_j + \sum\nolimits_{k=1}^{K} \hat{\alpha}_{jk} \ln \langle x_k \rangle_W .\end{aligned}$$

In Docket No. R90–1, UPS witness Nelson proposed this method using actual "costs" (i.e., real labor input) as weights, i.e., $\sigma_i = HRS_i \big/ \sum_{i=1}^{N} HRS_i$; Advo witness Lerner proposed using fitted costs as weights, i.e., $\sigma_i = \widehat{HRS}_i \big/ \sum_{i=1}^{N} \widehat{HRS}_i$.

<u>Comparison of arithmetic and geometric mean methods</u>

I computed elasticities for the ten cost pools with updated models, and the composite variability for those pools, using the arithmetic mean, geometric mean, and weighted geometric mean methods. The composite variability is the ratio of volume-variable costs to total costs for the cost pools. For the weighted geometric mean method, I used actual workhours as weights. In Table D–1, I present results based on the entire samples used to estimate the labor demand models. Table D–2 contains results based on the FY1998 subsets of the regression samples.

Table D–1. Cost pool and composite elasticities from alternative aggregation methods, using full regression sample observations

Cost Pool	Arithmetic mean method (USPS BY98)	Geometric mean method	Weighted geometric mean method
BCS	89.5%	91.2%	92.5%
OCR	75.1%	72.7%	74.3%
FSM	81.7%	82.4%	81.5%
LSM	95.4%	93.6%	95.5%
Manual Flats	77.2%	74.4%	74.5%
Manual Letters	73.5%	70.2%	68.8%
Manual Parcels	52.2%	50.8%	54.8%
Manual Priority	52.2%	55.7%	50.9%
SPBS	64.1%	61.9%	65.3%
Cancellation & Meter Prep	54.9%	55.6%	53.9%
Composite Variability	76.0%	75.1%	74.9%

Table D–2. Cost pool and composite elasticities from alternative aggregation methods, using FY1998 subset of regression sample observations

Cost Pool	Arithmetic mean method	Geometric mean method	Weighted geometric mean method
BCS	94.6%	94.6%	96.0%
OCR	84.1%	82.7%	84.3%
FSM	81.4%	82.4%	81.3%
LSM	96.6%	85.1%	96.4%
Manual Flats	79.0%	75.5%	75.6%
Manual Letters	68.9%	67.1%	66.2%
Manual Parcels	49.1%	47.6%	51.0%
Manual Priority	55.9%	59.7%	54.1%
SPBS	58.6%	56.7%	60.0%
Cancellation & Meter Prep	49.5%	50.3%	48.6%
Composite Variability	75.6%	74.8%	74.7%

Appendix E. Principal results from the "between" regression model

Table E–1. Principal results for letter and flat sorting operations, between model Uses full data set (FY1993 PQ2-FY1998 PQ4)

Cost Pool	BCS	OCR	FSM	LSM	Manual Flats	Manual Letters
Output Elasticity (Volume-variability factor)	1.044 (.067)	1.101 (.078)	1.026 (.049)	.913 (.046)	.963 (.084)	.906 (.074)
Deliveries Elasticity	.097 (.061)	-.065 (.070)	.051 (.041)	.077 (.037)	.172 (.070)	.104 (.067)
Wage Elasticity	-.229 (.219)	.372 (.306)	.007 (.237)	-.713 (.242)	.121 (.349)	-.059 (.341)
Capital Elasticity	-.003 (.043)	-.019 (.055)	-.023 (.036)	.048 (.029)	.063 (.058)	.132 (.046)
Manual Ratio Elasticity	.024 (.073)	-.067 (.088)	-.075 (.043)	-.077 (.045)	-.206 (.102)	-.215 (.082)
Adjusted R-squared	.963	.920	.967	.965	.947	.952
N observations	297	289	235	273	277	299
N sites	297	289	235	273	277	299

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.

1 **Table E–2. Principal results for other operations with piece handling data,**
2 **between model. Uses full data set (FY1993 PQ2-FY1998 PQ4).**

Cost Pool	Manual Parcels	Manual Priority	SPBS	Cancellation & Meter Prep
Output Elasticity (Volume-variability factor)	.730 (.067)	.748 (.058)	.889 (.116)	.845 (.076)
Deliveries Elasticity	.258 (.104)	-.011 (.110)	-.091 (.135)	-.007 (.072)
Wage Elasticity	1.007 (.744)	.227 (.638)	-.378 (.458)	.808 (.452)
Capital Elasticity	.149 (.080)	.334 (.088)	.056 (.090)	.135 (.051)
Adjusted R-squared	.778	.890	.728	.905
N observations	181	200	94	290
N sites	181	200	94	290

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.

1 **Table E–3. Principal results for letter and flat sorting operations, between**
2 **model. Uses data of "rate cycle" length (FY1996-FY1998)**

Cost Pool	BCS	OCR	FSM	LSM	Manual Flats	Manual Letters
Output Elasticity (Volume-variability factor)	1.060 (.070)	1.177 (.092)	1.043 (.053)	.873 (.076)	.911 (.085)	.933 (.072)
Deliveries Elasticity	.107 (.064)	-.112 (.079)	.068 (.045)	.135 (.072)	.162 (.074)	.092 (.064)
Wage Elasticity	-.137 (.219)	.292 (.359)	.185 (.231)	-.908 (.402)	-.094 (.384)	-.108 (.337)
Capital Elasticity	-.016 (.045)	-.038 (.063)	-.036 (.040)	.122 (.053)	.122 (.059)	.132 (.046)
Manual Ratio Elasticity	.036 (.071)	-.127 (.097)	-.028 (.047)	-.116 (.099)	-.163 (.103)	-.191 (.089)
Adjusted R-squared	.962	.909	.967	.958	.941	.950
N observations	282	260	227	116	223	270
N sites	282	260	227	116	223	270

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.

1 **Table E–4. Principal results for other operations with piece handling data,**
2 **between model. Uses data of "rate cycle" length (FY1996-FY1998)**

Cost Pool	Manual Parcels	Manual Priority	SPBS	Cancellation & Meter Prep
Output Elasticity (Volume-variability factor)	.769 (.075)	.751 (.062)	.827 (.113)	.901 (.084)
Deliveries Elasticity	.255 (.136)	-.052 (.118)	-.000[73] (.121)	-.039 (.079)
Wage Elasticity	.918 (.940)	.291 (.702)	-.436 (.449)	.568 (.479)
Capital Elasticity	.145 (.108)	.323 (.085)	.125 (.091)	.132 (.055)
Adjusted R-squared	.724	.887	.743	.901
N observations	139	171	93	270
N sites	139	171	93	270

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.

[73] Elasticity is –0.0004 (rounds to zero).

Appendix F. Principal results from the "pooled" regression model

Table F–1. Principal results for letter and flat sorting operations, pooled model

Cost Pool	BCS	OCR	FSM	LSM	Manual Flats	Manual Letters
Output Elasticity (Volume-variability factor)	.931 (.027)	.862 (.037)	.913 (.021)	.922 (.026)	.842 (.025)	.845 (.024)
Deliveries Elasticity	.207 (.032)	.161 (.045)	.148 (.023)	.065 (.033)	.281 (.031)	.223 (.029)
Wage Elasticity	-.754 (.051)	-.503 (.068)	-.653 (.042)	-.241 (.077)	-.280 (.056)	-.759 (.048)
Capital Elasticity	.041 (.016)	.011 (.024)	.043 (.012)	.046 (.019)	.095 (.018)	.082 (.015)
Manual Ratio Elasticity	.050 (.014)	-.013 (.019)	-.047 (.011)	-.045 (.017)	-.076 (.026)	-.184 (.019)
Auto-correlation coefficient	.884	.904	.891	.889	.907	.903
Adjusted R-squared	.975	.949	.990	.986	.978	.983
N observations	5390	5088	4357	3889	4879	5499
N sites	297	289	235	273	277	299

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.

Table F–2. Principal results for other operations with piece handling data, pooled model

Cost Pool	Manual Parcels	Manual Priority	SPBS	Cancellation & Meter Prep
Output Elasticity (Volume-variability factor)	.645 (.032)	.642 (.026)	.724 (.043)	.643 (.039)
Deliveries Elasticity	.246 (.056)	.254 (.057)	.004 (.058)	.318 (.046)
Wage Elasticity	-.656 (.142)	-1.016 (.144)	-1.201 (.078)	-.531 (.083)
Capital Elasticity	.173 (.037)	.254 (.039)	.094 (.031)	.084 (.019)
Autocorrelation coefficient	.882	.824	.899	.930
Adjusted R-squared	.891	.919	.977	.960
N observations	3023	3240	1569	5235
N sites	181	200	94	290

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.

Appendix G. Principal results from the "random effects" regression model

Table G–1. Principal results for letter and flat sorting operations, random effects model

Cost Pool	BCS	OCR	FSM	LSM	Manual Flats	Manual Letters
Output Elasticity (Volume-variability factor)	.916 (.030)	.821 (.041)	.880 (.024)	.918 (.027)	.802 (.028)	.790 (.026)
Deliveries Elasticity	.227 (.036)	.209 (.050)	.178 (.027)	.071 (.036)	.337 (.037)	.294 (.034)
Wage Elasticity	-.784 (.051)	-.542 (.069)	-.670 (.042)	-.236 (.078)	-.290 (.056)	-.775 (.048)
Capital Elasticity	.041 (.017)	.011 (.025)	.046 (.013)	.040 (.020)	.072 (.018)	.055 (.016)
Manual Ratio Elasticity	.053 (.014)	-.009 (.019)	-.047 (.011)	-.045 (.017)	-.040 (.026)	-.171 (.020)
Auto-correlation coefficient	.895	.924	.914	.909	.936	.932
Adjusted R-squared	.974	.942	.989	.985	.973	.979
N observations	5390	5088	4357	3889	4879	5499
N sites	297	289	235	273	277	299

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.

1 **Table G–2. Principal results for other operations with piece handling data,**
2 **random-effects model**

Cost Pool	Manual Parcels	Manual Priority	SPBS	Cancellation & Meter Prep
Output Elasticity (Volume-variability factor)	.615 (.036)	.627 (.029)	.662 (.049)	.569 (.043)
Deliveries Elasticity	.281 (.065)	.317 (.064)	.056 (.072)	.422 (.054)
Wage Elasticity	-.715 (.142)	-1.113 (.146)	-1.230 (.078)	-.580 (.083)
Capital Elasticity	.153 (.040)	.225 (.043)	.076 (.034)	.064 (.020)
Autocorrelation coefficient	.909	.853	.925	.953
Adjusted R-squared	.877	.914	.973	.949
N observations	3023	3240	1569	5235
N sites	181	200	94	290

Elasticities evaluated using arithmetic mean method; standard errors in parentheses.